The Role of Chromatin Structure in Regulating the Human Epidermal Differentiation Complex

Duncan Sproul

Thesis presented for the degree of Doctor of Philisophy,

University of Edinburgh 2008

Declaration

I declare that this thesis has been composed by me, and that all of the work is my own unless otherwise stated.

Duncan Sproul

November 2007

Acknowledgements

Firstly, I would like to thank Wendy for the opportunity to work on this project and for her supervision and enthusiasm over the three years. Nick has also been a great source of enthusiasm for the project and has always been available to answer questions and discuss results, even after he'd managed to escape across the road. The other members of the Bickmore lab (past and present): Shelagh, Liz, Heidi, Catherine, Emily, Inga, Anne, Steph, Celine, Clemence, Severine, Ragnhild, Annette and Elodie have all helped over the years, both by entertaining my sometimes ridiculous questions and for the unlimited supply of cake. Everyone else at the HGU also deserves thanks, but especially Paul Perry (Master of the Microscopes), Craig Nicol (for helping with the printing), Colin Semple and James Prendergast (for making up for my lack of computer knowledge), Stewart Mackay and the technical services team and Richard Baldock and Bill Hill for helping me understand Fourier transformations. This project also involved the help of a number of collaborators who I'd like to thank: Birgit Lane supplied me with Keratinocyte cell lines, Barbara Gorick and the Clone Service at the Sanger centre provided me with BACs and Fosmids, Claire Wallace, Donald Innes and Philippe Rigault helped design and order oligos for the arrays, Buake Ylstra supplied me with expression arrays and Margaret Hughes and the microarray facility at Liverpool printed arrays for me. Last but not least, I'd like to thank Anna and my parents for putting up with it.

Abstract

The Epidermal Differentiation Complex (EDC) is a co-ordinately regulated locus that is evolutionarily conserved within mammals. It consists of a large number of genes, organised into clusters of gene families, which mainly encode structural constituents of the cornified envelope which replaces the plasma membrane of fully differentiated keratinocytes. It is thought that the developmental program of gene expression at the locus is regulated by specific changes in chromatin structure (Williams *et al.*, 2002). To investigate this, I have characterised the chromatin structure of the EDC in cultured cell lines. These include a keratinocyte cell line, HaCaT cells, in which the locus is active and control cell lines where the locus is inactive. Chromatin is structured on a number of different levels, by the covalent modification of nucleosomes, the arrangement of nucleosomes into chromatin fibres and the arrangement of chromatin fibres into higher order structures within the interphase nucleus. I have assayed chromatin structure on all these levels using Chromatin Immunoprecipitation and Sucrose Gradient Sedimentation Analysis of Chromatin Fibre structure, partnered with oligonucleotide microarrays and Fluorescent *In-Situ* Hybridisation. By doing so I have examined the role each level of chromatin structure plays in regulating the human EDC and, characterised the relationships between the different levels across a large co-ordinately regulated locus in the human genome.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3C – Chromosome Conformation Capture

4C – Chromosome Conformation Capture on ChIP

aCGH – Array comparative genome hybridisiation

BAC- Bacterial Artificial Chromosome

bp- basepairs

BPC- Buffered Phenol-Chloroform

BSA- Bovine Serum Albumin

chr- chromosome

ChIP – Chromatin Immunoprecipitation

ChIP-chip – ChIP on chip (ChIP combined with microarrays)

CK10- Cytokeratin 10

CK14- Cytokeratin 14

CO_2- Carbon Dioxide

CCCTC-binding factor- CTCF

Cy3- Cyanine 3

Cy5- Cyanine 5

C-terimus- Carboxyl-terminus

d^2- Mean Squared Interphase Separation of Probes in the Nucleus

Da- Daltons

DAPI- 4,6-diamino-phenylindole

DFT- Discrete Fourier Transformation

DMEM- Dulbecco's Modified Eagle's Medium

DMSO- Dimethyl Sulfoxide

DNA- Deoxyribonucleic acid

EDC- Epidermal Differentiation Complex

EDTA- Ethylenedinitrilotetraacetic acid

ENCODE- Encyclopedia of DNA Elements

FCS- Foetal Calf Serum

FFT- Fast Fourier Transformation (computational algorithm for performing DFTs)

FITC- Fluorescein isothiocyanate

FLG- Filaggrin

FISH– Fluorescent *in situ* hybridisation

GCR- Global Control Region

GEO- Gene Expression Omnibus

H3K9- Histone H3 Lysine 9

H3K27- Histone H3 Lysine 27

H4K16- Histone H4 Lysine 16

HAT- Histone Acetyltransferase

HDAC- Histone Deacetylase

HCl- Hydrochloric Acid

hr- hour

HS- Hypersensitive Site

IVL- Involucrin

KS- Test: Kolmogorov-Smirnov test

LCE- Late Cornified Envelope

MAA- Methanol Acetic Acid

MHC-Major Histocompatibility Complex

Min- minute

MNase- Micrococcal Nuclease

MOPS- 3-(N-Morpholino)propanesulfonic acid

Na Butyrate- Sodium Butyrate

NaCl- Sodium Chloride

NaOH- Sodium Hydroxide

NB- Nuclear Buffer

NChIP- Native Chromatin Immunoprecipitation

N-Terminus- Amino Terminus

Oligo – oligonucleotide

OMIM- Online Mendelian Inheritance in Man

OR- Olfactory Receptors

PAC- P1 Artificial Chromosome

PBS- Phosphate Buffered Saline

PCR – polymerase chain reaction

PRC- Polycomb Recongnition Compex

RNA- Ribonucleic acid

RT- Room Temperature

RT-PCR - Reverse transcription polymerase chain reaction

SDS- Sodium Dodecyl Sulfate

SDS-PAGE- Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEM- Standard Error in the Mean

SPRR- Small Proline Rich

SSC- Saline Sodium Citrate

TCA- Trichloroacetic Acid

TEMED- N,N,N',N'-Tetramethylethylenediamine

TSA- Trichostatin A

U- Units

UV- Ultraviolet

WGA- Whole Genome Amplification, Kit from Sigma

Chapter 1: Introduction

The DNA sequence is only the first step in understanding the human genome. It is the differential expression of genes in time and space that ultimately provides the genome's function. 3.2 billion basepairs (bp) of DNA measuring roughly 2m are compacted with proteins to fit within in the confines of a cell's nucleus with an average volume of around 1500μm^3. Cellular DNA in complex with its packaging proteins is known collectively as chromatin and serves as the primary physiological substrate for DNA-dependent pathways such as transcription, replication and repair. Therefore, the organisation of DNA in the mammalian nucleus is an integral component in the regulation of gene expression. How does chromatin package DNA within the nucleus and how does chromatin regulate gene expression?

1.1 Primary and Secondary Chromatin Structure

1.1.1 Histones and Nucleosomes

The most abundant constituents of chromatin are histones, which are small (11-16KDa), basic proteins. They are among the most evolutionarily conserved of all proteins, but not all organisms use them to package DNA. Prokaryotes appear to have evolved a variety of different, phylogenetically unrelated proteins to fulfil this role, and those found in Archea are thought to be the evolutionary ancestors of eukaryotic histones (Sandman et al., 1998). Even among the eukaryotes, the dinoflagellates are known to package their DNA with small basic proteins that are unrelated to histones (Vernet et al., 1990) and in mammalian spermatozoa DNA is compacted using proteins known as protamines (Wolffe, 1998).

In most eukaryotes the majority of histones are of four types collectively known as the core histones, Histones H2A, H2B, H3 and H4. They form an octamer complex *in vivo* together with 100 to 170bp of DNA to give a structure known as the

nucleosome (van Holde and Zlatanova, 1999). Histones are so abundant in the cell that they are easily visible on coomassie stained SDS PAGE gels of cellular extracts and they are encoded in the human and mouse genome by genes arranged in several large clusters (Marzluff et al., 2002). The core histones have two structural domains, a histone fold domain which interacts with the DNA through positively charged residues (Arents et al., 1991;Luger et al., 1997) and a N-terminal domain which is thought to be a flexible, tail like structure which protrudes from the core of the nucleosome (Schroth et al., 1990).

1.1.2 Nucleosome Positioning

Nucleosomes were determined to be the fundamental, repeating unit of chromatin through a series of experiments utilising both nuclease digestion and electron microscopy. Firstly, nuclease digestion showed that chromatin was degraded in a series of steps of roughly 200bp (Hewish and Burgoyne, 1973). By extensive digestion, it was discovered that these independent 200bp units consisted of DNA in a complex with proteins (Rill and Van Holde, 1973;Sahasrabuddhe and Van Holde, 1974). Finally electron micrographs of chromatin prepared under certain conditions showed chromatin to be made up of nucleosome 'beads' joined by 'strings' of DNA (Oudet et al., 1975).

Classical experiments demonstrated that the packaging of DNA in nucleosomes serves to inhibit both prokaryotic and eukaryotic RNA polymerases (Allfrey et al., 1963;Huang and Bonner, 1962;Knezetic and Luse, 1986;Lorch et al., 1987). It could be imagined that if a particular DNA sequence was wrapped around a nucleosome, then it would be inaccessible to other DNA binding proteins such as transcription factors, particularly because histones are known to have a high affinity for DNA (Burton et al., 1978).

Histones recognise the minor groove of DNA rather than the bases and so bind to DNA non-specifically. Nevertheless it is known that certain DNA sequences preferentially bind nucleosomes (Luger et al., 1997). Due to the need of DNA to bend when wrapped around a nucleosome, stiff DNA sequences such as poly-A stretches are poorly incorporated into nucleosomes whereas flexible sequences

containing AT dinucleotides with a 10-bp periodicity have high affinities for nucleosomes (Prunell, 1982;Widlund et al., 1997).

On a naked DNA template, these factors are very important for nucleosome positioning, but it is unclear as yet how much they influence chromatin structure *in vivo*. It was found that the nucleosomal arrangement of a selection of yeast promoters could be recapitulated *in vitro* solely from DNA and histones (Sekinger et al., 2005). Preferential dinucleotide binding motifs are also found with a non-random periodicity in the genomes of some eukaryotes (Widom, 1996) and binding sites are regularly spaced across the chicken *β-Globin* locus (Davey et al., 1995). However, the overall nucleosome association properties of the chicken genome were found to be very similar to that of random DNA implying that sequences with a high nucleosome affinity are found in less than 5% of the genome (Lowary and Widom, 1997). Recent studies have used microarrays to examine nucleosome positioning genome wide in yeast and agree that the majority of nucleosomes, at least in this relatively compact genome, are stably positioned (Segal et al., 2006;Yuan et al., 2005;Peckham et al., 2007). However, they disagree as to whether some kind of histone positioning code exists in the genome sequence or not.

Despite the apparent stability of many nucleosome positions, it is known that their mobility of nucleosomes is a key factor in gene expression (Meersseman et al., 1992). Many different ATP dependent chromatin remodelling enzymes exist in the cell. Most of these enzymes fall into three different classes depending on their domain structure. Different enzymes act on nucleosomes in different ways but there appears to be a large degree of overlap and the exact mode of action of different enzymes remains under debate (for a review see, de la Serna et al., 2006).

1.1.3 Chromatin Fibres

The positioning of nucleosomes on the DNA sequence represents the primary level of chromatin structure, but nucleosomes do not bind to DNA in isolation within the nucleus. It can be imagined that the binding of one nucleosome to a strong site could sterically hinder the binding of a second nucleosome to close but slightly weaker site, this is known as ‘packing effect’ (Rando and Ahmad, 2007).

Furthermore, *in vivo* it is not believed that the 'beads on a string' arrangement of chromatin occurs. Instead nucleosomes are packed into chromatin fibres which are thought to have a diameter of ~30nm and are therefore known as 30nm fibres. Early electron micrographs showed that the appearance of chromatin fibres isolated from cells could be varied depending on the salt concentration (Oudet et al., 1975;Thoma et al., 1979). It was hypothesised that conformational differences such as these may occur locally in the nucleus and so regulated transcription. More compact structures would be refractory to transcription and open structures permissive.

The exact structure that the 30nm fibre adopts *in vivo* remains a point of debate. One proposed model is the solenoid model, in which nucleosomes are connected by bent linker DNA and are arranged linearly into a higher order helix (Finch and Klug, 1976). The second major model is known as the 'zig-zag' model with straight linker DNA that zig-zags back and forth between adjacent helical stacks of nucleosomes (Woodcock et al., 1984). Over the years, various experiments have been interpreted as supporting the existence of one or the other of these models (for examples of reviews drawing different conclusions see, (Horowitz-Scherer and Woodcock, 2006;Robinson and Rhodes, 2006;Gilbert et al., 2005).

Recently, the X-ray structure of a tetranucleosome was solved to a 9 Angstrom resolution. It showed a zig-zag model where the two stacks of nucleosomes are not parallel but rotated at an angle of –71.3° with respect to each other (Schalch et al., 2005). However, the repeat length (amount of DNA per nucleosome, (Wolffe, 1998)) of the fibre analysed was 167bp and the authors estimated the structure could tolerate a variation in this value of +/-5bp. This falls short of the repeat length more usually found in mammalian cells of around 180-200bp (Sumner, 2003).

In vivo chromatin is not constituted onto regular templates, imaging and modelling of such chromatin structures has concluded that they form irregular chromatin fibres (Woodcock et al., 1993;Woodcock, 1994;Horowitz et al., 1994;Leuba et al., 1994). Biophysical analysis of mouse satellite repeats showed that they have a more compact chromatin structure than the bulk of the genome (Gilbert and Allan, 2001). This is likely to be because their repetitive DNA sequence allows a very regular spacing of nucleosomes and an ordered and compact chromatin fibre.

Modelling of the biophysical properties of bulk chromatin fibres on the other hand suggests that they have a less regular conformation and are interspersed with discontinuities, probably generated by irregular nucleosome spacing and the binding of other factors to the DNA (Caplan et al., 1987;Gilbert and Allan, 2001). Finally, a detailed analysis of an inactive region upstream of the chicken β-globin found that it had biophysical properties consistant with a compact rod-like 30nm chromatin fibre (Ghirlando et al., 2004).

1.2 Factors Influencing Primary and Secondary Chromatin Structure

1.2.1 Linker Histones

As well as the 4 core histones, another type of histone is found in cells. These are linker histones which consist of a central globular domain flanked by N- and C-terminal domains (Aviles et al., 1978;Hartman et al., 1977). Linker histones do not form part of the core nucleosome but instead bind to the outside of this structure by the interaction of their globular domains with both DNA and the core histone H2A (Boulikas et al., 1980;Allan et al., 1980). The structure of the globular domain has been solved and it shows a high degree of similarity to the DNA binding proteins CAP and RAP30 as well as the transcription factor HNF-3 (Ramakrishnan et al., 1993;Clark et al., 1993;McKay and Steitz, 1981;Groft et al., 1998). The globular domain also has the ability to bind to naked DNA (Thomas et al., 1992;Draves et al., 1992).

Many *in vitro* studies have suggested different models for the binding of the linker histone to the DNA and nucleosome (Travers, 1999). Linker histones *in vivo* are known to be very mobile (Misteli et al., 2000), and may occupy different positions in relation to the nucleosome depending on its exact context. A study of fluorescently tagged linker histone mutants has defined that the bulk of linker histones in a cell bind DNA in two separate locations, firstly at the nucleosome dyad

on the nucleosomal surface and secondly to the minor groove of linker DNA between two nucleosomes (Brown et al., 2006a). Both the C-terminal tail and the globular domain of the linker histone are necessary for this binding.

Experiments into the folding of chromatin fibres have shown that the presence of the linker histone can compact the chromatin fibre (Thoma et al., 1979;Thoma and Koller, 1977). Comparisons of chromatin folding under ionic conditions in both the presence and absence of linker histone have suggested that both the linker histones and the core histone tails contact the linker DNA between nucleosomes to facilitate the compaction of nucleosomal arrays (Clark and Kimura, 1990;Fletcher and Hansen, 1995;Garcia-Ramirez et al., 1992;Garcia-Ramirez et al., 1995;Schwarz et al., 1996). This suggests that the neutralisation of charges is likely to be an important factor in the folding of nucleosomes into higher order structures (Wolffe, 1998).

Most organisms don't just have one linker histone type, different linker histone genes encode proteins with different structures and are expressed in a highly regulated manner depending on the stage of the cell cycle, tissue type and developmental stage (Lennox and Cohen, 1983;Marzluff et al., 2002;Wang et al., 1997). In mice, the complement of histone H1 genes includes at least 8 members, the somatic variants *H1a* to *e*, two germ cell specific forms, *H1t* and *H1oo*. A third variant is the replacement linker histone $H1^0$, which is synthesised in a replication independent manner and accumulates in quiescent cells (reviewed in (Zlatanova and Doenecke, 1994).

A series of mouse knockout studies have been performed to examine linker histone function. Firstly, a mouse lacking the $H1^0$ gene developed normally because other linker histone types were upregulated in its absence (Sirotkin et al., 1995). In addition, any one of three of the *H1c*, *d* or *e* subtypes was found to be dispensable even in $H1^0$ knockout mice (Fan et al., 2001). The ablation of all three of these subtypes is lethal but, just a single functional allele of one of these three was sufficient to partially rescue the embryonic lethality, albeit with defects in growth and altered nucleosomal spacing (Fan et al., 2003). Embryonic stem cells generated from the triple-null embryos propagate normally but have an altered chromatin structure (Fan et al., 2005). Although linker histones are though to play a general

role in chromatin structure, these knockout cells show very specific changes in gene expression. The *H1b* subtype has also been seen to have specific role in the regulation of the *MyoD* gene during muscle differentiation (Lee et al., 2004).

1.2.2 Other Chromatin Proteins

Another important group of chromatin proteins are the High Mobility Group (HMG) proteins, so called because of their small size and mobility on electrophoretic gels (Goodwin et al., 1973). There are a number of different classes of HMG proteins but their total complement in cells is thought to add up to somewhat less than the number of nucleosomes in a cell, so they are less abundant than linker histones (Seale et al., 1983). One of these, the HMGB1 protein has been seen to disassociate from metaphase chromosomes, implying that HMG proteins are also a less stable component of the chromatin fibre than linker histones (Falciola et al., 1997). Comparisons of the FRAP kinetics of tagged HMG-17 and 14 with those of linker histones H1o and H1c suggest that HMG proteins exchange more rapidly on chromatin than linker histones (Misteli et al., 2000;Phair and Misteli, 2000).

In general, HMG proteins are thought to destabilise compacted chromatin structures and promote a more 'open' chromatin fibre (Ding et al., 1997). This means that they could be considered, transcriptional activators rather than repressors. Despite this, it is known that one class, HMGI, binds to satellite repeat regions of the genome which are though to have a condensed chromatin structure (Radic et al., 1992). They have been shown to play a role alongside some ATP dependent chromatin remodelling complexes, perhaps facilitating their access to the DNA (Bonaldi et al., 2002). Early binding data suggested that perhaps linker histones and HMG group proteins competed for binding sites on the chromatin fibre to set up differing fibre structures (Hill and Reeves, 1997;Zlatanova and van Holde, 1998). FRAP has been used to show that HMG proteins from 3 different groups are capable of competing for distinct binding sites with Histone H1 but, whereas members of the same HMG family compete for binding sites, there is no competition between different families (Catez et al., 2002;Catez et al., 2004).

Cells lacking the HMG-1 protein are viable and don't display any general defects in chromatin structure but misexpress the glucocorticoid receptor (Calogero et al., 1999). A number of HMG proteins have been seen to interact with specific transcription factors (Jayaraman et al., 1998;Zappavigna et al., 1996;Zwilling et al., 1995). This suggests that although they are general chromatin proteins, different HMG proteins may act to regulate specific subsets of genes, like the linker histones.

Polycomb group genes were first identified in *Drosophila* as mutants that affected the regulation of homeotic genes and encode chromatin associated proteins (Zink and Paro, 1995). These proteins form two complexes in mammals, Polycomb Repressive Complex 1 (PRC1, Shao et al., 1999) and Polycomb Repressive Complex 2 (PRC2, Kuzmichev et al., 2002). Members of the PRC2 complex are stably associated with the inactive X in female mouse cells and are important for its inactivation (Mak et al., 2002;Wang et al., 2001). An association with transcriptionally repressed chromatin would imply that these proteins may compact chromatin. Using electron microscopy it was shown that the PRC1 complex could indeed compact short arrays of nucleosomes (Francis et al., 2004). This process was mediated by the protein PCC and did not require the N-terminal tails of the histones. The same complex is capable of blocking chromatin remodelling by the SWI/SNF ATP dependent chromatin remodelling complex (Shao et al., 1999). Another chromatin associated protein PC4 is also capable of compacting nucleosome arrays but does so in a different manner to linker histones (Das et al., 2006).

1.2.3 Core Histone Variants

With the exception of Histone H4, all of the core histones have a number of variants which probably arose through gene duplication (Malik and Henikoff, 2003). The best characterised histone variant is CENP-A which has a highly specialised role in the cell. It is associated with chromosomal kinetochores and is vital for their function (Palmer et al., 1991;Stoler et al., 1995). CENP-A is targeted to the centromere by its histone fold domain (Sullivan et al., 1994). The fact that a histone variant is targeted to a specific functional part of the chromosome gives rise to the idea that histone variants set up specific functional regions or domains of chromatin.

The canonical core histones are encoded in specialised RNAs which lack a polyA tail (Dominski and Marzluff, 1999). Their synthesis and deposition onto DNA is tightly regulated and occurs only during S-phase as new DNA strands are synthesised (Borun et al., 1975;Mello and Almouzni, 2001). Histone variants are often encoded by genes with a more normal structure which are often synthesised and deposited onto DNA outside of S-phase (Borun et al., 1977;Wu and Bonner, 1981;Wu et al., 1982).

1.2.3.1 Variants of H3

Besides CENP-A, three further versions of Histone H3 are found in mammals, H3.1, H3.2 and H3.3 which all have very similar amino acid sequences (Hake and Allis, 2006). Despite their similarities, it was suggested they may be responsible for mediating different chromatin fibre structures (Franklin and Zweidler, 1977). CENP-A has a protein structure that differs significantly from the canonical Histone H3 protein and it has been shown that the incorporation of this variant into chromatin leads to the formation of a more rigid, compact fibre (Black et al., 2004). Recently, it has also been shown that in *Drosophila*, CENP-A containing nucleosomes are made up of a histone tetramer rather than an octamer and the authors believe this may also be the case in other organisms (Dalal et al., 2007a;Dalal et al., 2007b). H3.1 and H3.2 are both expressed in a replication dependent manner and only differ from each other by a single amino acid. They are generally considered to be functionally identical in most studies but recently it has been suggested that H3.2 is associated with transcriptionally inactive chromatin (Hake et al., 2006). However, no direct structural effect of the incorporation of H3.2 into chromatin has been described.

H3.3 differs from H3.1 at four amino acid residues and is known to be incorporated into chromatin in a replication independent fashion (Ahmad and Henikoff, 2002). Three of these amino acid substitutions have been shown to be required for replication independent incorporation into chromatin (Ahmad and Henikoff, 2002) and a specific complex for the assembly of H3.3 into chromatin has been purified (Tagami et al., 2004). The equivalent region of CENP-A also confers

centromeric targeting to a histone H3 hybrid (Black et al., 2004) and the sole amino acid difference between H3.1 and H3.2 also occurs in this region but it is unknown whether H3.2 is incorporated by a separate complex (Hake and Allis, 2006).

The fact that H3.3 was seen to be incorporated into active rDNA arrays led to the suggestion that perhaps the variant marks transcriptionally active chromatin (Ahmad and Henikoff, 2002). To examine H3.3, many groups have made use of tagged versions of the protein to perform Chromatin Immunoprecipitation (ChIP) and shown that it is generally present at actively transcribed regions (Chow et al., 2005;Mito et al., 2005;Wirbelauer et al., 2005). In addition, fluorescently labelled H3.3 has been seen to be incorporated *in vivo* into a transgene array as it undergoes activation (Janicki et al., 2004). These associations with active genes would suggest that H3.3 conferred a different chromatin structure to these regions, but don't show that this is the case. Recently it was shown that nucleosomes containing tagged H3.3 are more sensitive to salt based disruption than those containing H3.2 (Jin and Felsenfeld, 2007), chickens do not contain H3.1 and the variant commonly referred to as H3 is equivalent to mammalian H3.2, (Hake and Allis, 2006).

1.2.3.2 Variants of H2A

The other commonly studied histone variants present in mammals are versions of H2A. One variant, macro-H2A histone variant is associated with the inactive X-chromosome in mammalian cells (Costanzi and Pehrson, 1998). The role of H2AX also appears to be highly specialised, it is associated with DNA repair and recombination (Malik and Henikoff, 2003). Another H2A variant, H2ABbd (Bbd-Barr Body Deficient), has a mutually exclusive localisation pattern compared to macro-H2A suggesting it may be present in transcriptionally active regions (Chadwick and Willard, 2001). Sedimentation analysis in different salt concentrations suggested that H2ABbd containing nucleosomes may be more unstable than those containing canonical histones (Gautier et al., 2004). These authors also showed by FRAP that GFP tagged H2ABbd exchanges more rapidly in the cell than similarly tagged H2A.

Another H2A variant, H2AZ, has been more widely studied and is essential in both mammals and in *Drosophila* (Faast et al., 2001;Clarkson et al., 1999). In budding yeast it is not essential but has a function that can not be substituted for by canonical H2A (Jackson and Gorovsky, 2000). It was found to be associated with the transcriptionally active macronucleus of *Tetrahymena thermophila*, leading to the suggestion that it is deposited in transcriptionally active chromatin (Allis and Wiggins, 1984). In yeast, H2AZ was found it to be associated with gene promoters and may also be positioned at the boundaries between different chromatin domains (Meneghini et al., 2003;Li et al., 2005;Zhang et al., 2005). Related patterns have been seen in metazoan organisms (Farris et al., 2005;Bruce et al., 2005;Swaminathan et al., 2005).

Studies of the properties of H2AZ containing nucleosomes have come to conflicting conclusions. The crystal structure of a H2AZ containing nucleosome suggested that its incorporation may subtly destabilise the nucleosome (Suto et al., 2000). Chicken nucleosomes reconstituted with H2AZ were also less stable when assessed by sedimentation at different ionic concentrations (Abbott et al., 2001). However, a separate series of studies have concluded that the incorporation of H2AZ into chromatin leads to more stable nucleosomes and more compact nucleosomal arrays (Fan et al., 2004;Fan et al., 2002;Park et al., 2004;Thambirajah et al., 2006). The ability of H2AZ to generate compact chromatin fibres may be mediated through the characteristics of an acidic patch located on the surface of the nucleosome which is contacted by the N-terminal tail of histone H4 on an adjacent nucleosome (Luger et al., 1997). In H2AZ it has a more acidic nature (Suto et al., 2000) but in H2ABbd it is less so (Chadwick and Willard, 2001). A recent study has suggested one reason why studies of H2AZ have come to such contrasting conclusions. They found that H3.2-H2A.Z containing nucleosomes were at least as stable as H3.2-H2A nucleosomes, but that the presence of H3.3 destabilised the nucleosome and that the partnership of H3.3 and H2AZ was even more unstable (Jin and Felsenfeld, 2007).

1.2.4 Histone Modifications

Another factor widely believed to modulate chromatin structure are histone modifications. The first modifications to be described were histone acetylation and methylation (Allfrey et al., 1964). The majority, of these modifications occur on the N-terminal tails of the core histones (*Figure 1.1*). There are two hypotheses regarding the role of histone modifications in regulating chromatin structure. In the first they are directly responsible for the alteration of chromatin structure. In the second they do not directly alter chromatin structure themselves but are recognised by other proteins which dock to them and carry out the alteration of chromatin structure. The idea of a 'histone' code was postulated, whereby the histone modifications provide a code that can be read to set up the chromatin landscape (Turner et al., 1992;Turner, 1993). Recently, a number of studies have examined the distribution of histone modifications across a large portion of the genome using microarrays in an effort to 'crack' the histone code (Birney et al., 2007). Furthermore, the availability of mass sequencing technology has enabled such studies to be conducted on a genome wide scale (Barski et al., 2007). Covalent modifications are also known to occur on other chromatin associated proteins such as Histone H1 (Wisniewski et al., 2007), and the HMG proteins (Sterner et al., 1979).

1.2.4.1 Histone Acetylation

Acetylation was one of the first histone medications to be discovered, and from the start it was associated with active transcription (Allfrey et al., 1964). It is known that the acetylation of histones is required for the process of nucleosome assembly (Mello and Almouzni, 2001) but it also occurs once they have become incorporated into chromatin. The reversible acetylation of nucleosomes present in the chromatin fibre is the subject of a great deal of interest with regard to chromatin structure, mainly because the acetylation of histone tails alters their charge, making them more negative (McGhee and Felsenfeld, 1980;Wolffe, 1998). Therefore, they may be able to directly alter chromatin structure without the need to bind accessory proteins.

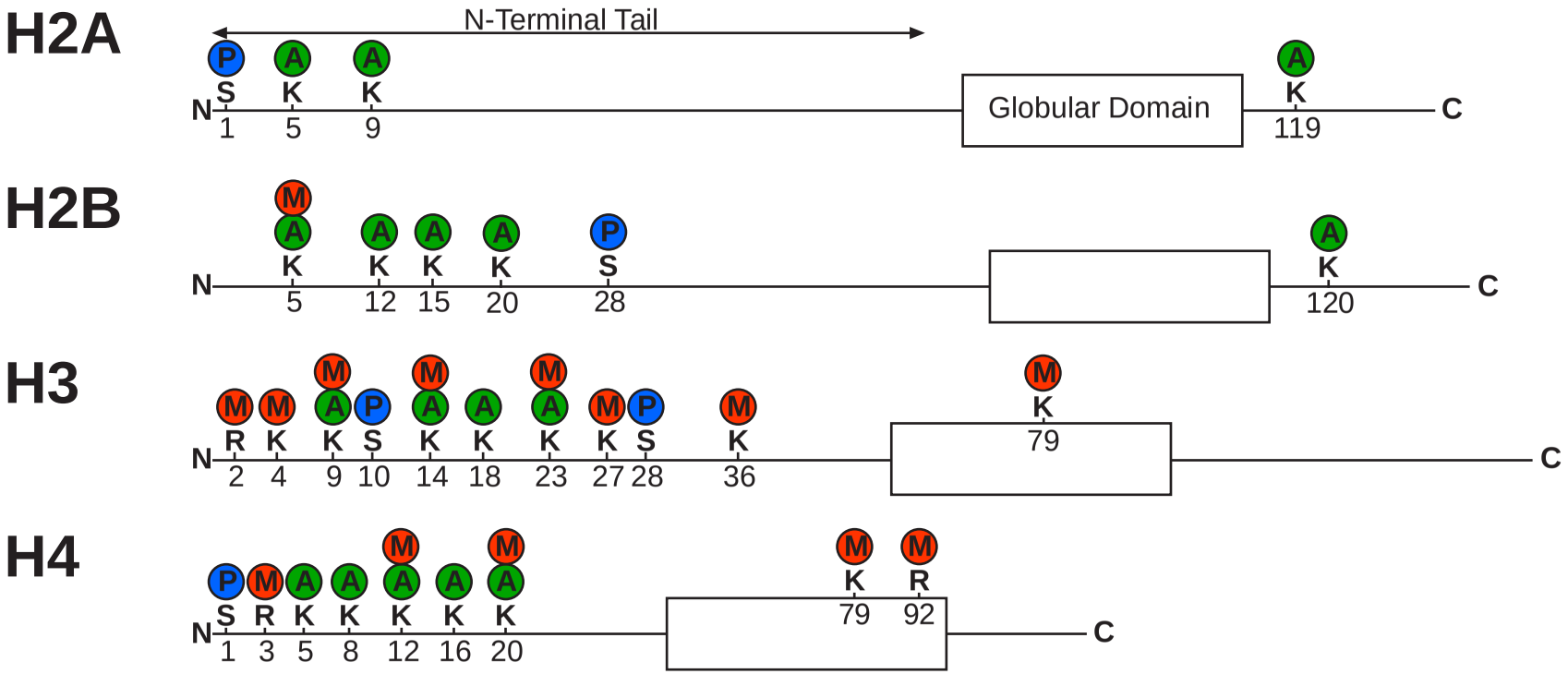


Figure 1.1 Modifications of the Core Histones

Specific residues of the core histones can be chemically modified by methylation (M), acetylation (A) or phosphorylation (P). Modifications other than those shown can occur (Kouzarides, 2007). Standard single letter amino acid codes are used to indicate the modified residues (ie K=Lysine, S=Serine and R=Arginine). The numbers below the residues indicate their exact positions within the primary sequence of the histone protein.

An early study used Chromatin Immunoprecipitation (ChIP) to show that the promoters of transcriptionally active genes are associated with acetylated histones (Hebbes et al., 1988). This conclusion has been supported more recent large scale ChIP studies that make use of microarrays or mass sequencing to gain information on a genome wide scale (Barski et al., 2007;Heintzman et al., 2007;Koch et al., 2007;Roh et al., 2005).

Given that an open chromatin structure is thought to be associated with transcription, this association suggests that acetylation of histones could modify chromatin structure. The effect of acetylation on the structure of single nucleosomes was found to be rather subtle in *in vitro* studies (Bode et al., 1983;Oliva et al., 1990). However, studies of the properties of acetylated chromatin fibres have suggested that histone acteylation inhibits the formation of a compact structure, at least *in vitro* (Garcia-Ramirez et al., 1995;Tse et al., 1998).

The discovery that transcriptional co-activators were enzymes capable of catalysing the acetylation of histones, further cemented the association between histone acetylation and transcription (Brownell et al., 1996;Ogryzko et al., 1996;Kuo et al., 1998). Another set of enzymes, Histone Deacetylases (HDACs) were discovered that catalyse the reverse reaction (Taunton et al., 1996;Hassig and Schreiber, 1997;Hassig et al., 1997). The dynamic interaction of the two processes is important for the cell as amphibian embryos grown in the presence of histone deacetylase inhibitors have their development severely disrupted (Almouzni et al., 1994). A specific protein domain, the bromodomain, has also been described that recognises and binds to acetylated histones (Dhalluin et al., 1999). This has meant that most recent research has been focused on the histone code hypothesis, ie which enzymes recognise and read acetylation marks.

One study has however addressed whether acetylation functions directly or via the recruitment of other proteins in yeast. Yeast mutants of histone H4 were generated where single lysine residues (lysines 5, 8, 12 and 16) had been replaced by arginine to render them unacetylable and encoding a permanent positively charged state (Dion et al., 2005). The four mutants were also combined in every possible permutation and the gene expression of each strain analysed using microarrays. The authors concluded that in general H4 acetylation acted in a simple combinatorial

manner, but that a number of yeast genes appeared to be specifically regulated by the acetylation of Histone H4 lysine 16 (H4K16Ac).

Another recent study has focused on a role of H4K16Ac in directly altering chromatin structure. The H4 tail, particularly between residues 14 and 23 makes inter-nucleosomal contacts and appears to be important in the formation of chromatin fibres (Dorigo et al., 2003;Luger et al., 1997). It was found that the acetylation of H4K16 alone was capable of interfere with the formation of compact chromatin fibres and inhibit the ATP dependent remodelling enzyme ACF from mobilizing a nucleosome (Shogren-Knaak et al., 2006). These two studies, combined with observation that H4K16Ac has a special role in sex chromosome dosage compensation in *Drosophila*, suggest the modification may have a special role in chromatin structure (Turner et al., 1992). However, at present, no association between histone acetylation and an open chromatin structure has been described *in vivo*.

1.2.4.2 Histone Methylation

Many lysine and arginine residues of the core histones are methylated and this can occur in mono-, di- and tri- methylated forms (Bannister et al., 2002;Zhang and Reinberg, 2001). Unlike acetylation, methylation does not alter the charge of lysine residues and therefore has not been proposed to directly modulate chromatin structure. The development of antibodies capable of differentiating the various mono-, di- and tri- methylated residues (Perez-Burgos et al., 2004) has led to an explosion of ChIP-based studies which have examined the association of these modifications with different genomic features.

Initial evidence associated the methylation of Histone H3 lysine 4 (H3K4) with histone acetylation and transcriptional activity. It was seen that an H3K4 lysine methyltransferase activity was associated with the transcriptionally active macronucleus of *Tetrahymena* (Strahl et al., 1999). A variety of localised ChIP studies suggested that indeed H3K4 methylation was indeed found at active gene promoters (Liang et al., 2004;Santos-Rosa et al., 2002;Schneider et al., 2004;Bulger et al., 2003). This association has been confirmed by a large scale ChIP studies in

many organisms (Barski et al., 2007;Bernstein et al., 2005;Heintzman et al., 2007;Koch et al., 2007;Bernstein et al., 2002;Schubeler et al., 2004). One study has suggested that the H3K4 methylation patterns can be used to distinguish different types of regulatory elements, active enhancers are marked with mono- and di- H3K4 methylation whereas active promoters are mainly marked by di- and tri- methylation (Heintzman et al., 2007). Recently it was shown that the TAF3 component of TFIID recognises and binds to H3K4me3 marks at promoters, directly linking the core transcriptional machinery to this histone modification (Vermeulen et al., 2007). In yeast a single enzyme, Set1, is responsible for all H3K4 methylation but in mammals a number of different enzymes exist which may each regulate different sets of genes (Ruthenburg et al., 2007).

Methylation of Histone H3 Lysine 9 (H3K9), is regarded as a repressive modification. Su(var)3-9 was identified as a H3K9 methyltransferase associated with inactive genomic regions and it was determined that H3K9me3 created a binding site for another protein associated with inactive regions of the genome, HP1 (Rea et al., 2000;Lachner et al., 2001;Bannister et al., 2001). The first ChIP studies of the modification also suggested an association with inactive regions (Litt et al., 2001). In general large scale studies of histone modifications have confirmed this hypothesis by showing that H3K9me2 and H3K9me3 are associated with the promoters of inactive genes, but surprisingly it was found that H3K9me1 is better associated with active transcriptional start sites (Barski et al., 2007).

Another methylation mark often associated with transcriptional repression is that of H3 lysine 27. The PRC2 complex contains the enzyme EED-EZH2 which has H3K27 methyltransferase activity and is involved in *Hox* gene silencing (Cao et al., 2002). This association has been confirmed by large scale ChIP studies (Barski et al., 2007;Bernstein et al., 2006). However, these studies have also found evidence for bivalent chromatin marked by both H3K4 and H3K27 methylation, which they suggest may mark genes poised for a transcriptional choice in ES cells (Bernstein et al., 2006). Such bivalent chromatin marks have also been observed in differentiated cells (Barski et al., 2007;Roh et al., 2006). Again, there have been no studies which connect the presence of histone methylation to a particular chromatin structure. It is known that mouse minor satellite sequences are marked by H3K9me3 and have a

compacted 30nm chromatin structure (Gilbert and Allan, 2001;Martens et al., 2005) but it has not been shown that the presence of the mark leads to the condensed chromatin state. As well as methylating H3K27, PRC complexes bind this modification and, as mentioned above, PRC1 has the ability to compact nucleosomal arrays (Francis et al., 2004).

1.2.4.3 Other Histone Modifications

Other lysines are known to be modified *in vivo,* including lysines 36 and 79 of Histone H3, lysine 20 of Histone H4 and lysine 5 of Histone H2B as well as various arginine residues of the core histones (Kouzarides, 2007). However, acetylation and methylation are not the only modifications that are present on histones. Another highly studied modification is histone phosphorylation which has strongly been associated with either cell cycle progression (Wei et al., 1999) or DNA repair (van Attikum and Gasser, 2005). Mass spectrometry studies have provided evidence for the existence of a large number of different histone modifications such as ubiquitination, sumoylation, ADP ribosylation, deimination and proline isomerization (Bonaldi et al., 2004). However at present not a great deal is known about their role in the cell and their ability to modulate chromatin structure (Kouzarides, 2007).

1.2.4.4 Stability of Histone Modifications

In considering histone modifications, it is worth remembering that they are dynamic and the states measured by ChIP represent their steady state levels. Analyses using HDAC inhibitors show that treatment with these drugs quickly results in hyperacetylation of a significant portion of histones and their removal leads to deacetylation (Spencer and Davie, 2001). Therefore, the action of both histone acetyltransferases and HDACs constantly affects a pool of histones in the cell. When inhibitors are removed, two distinct groups of acetylated histones are observed. The first consists of hyperacetylated histones (ie tri or tetra acetylated histones) that are

only seen upon exposure to HDAC inhibitors, termed class I acetylated histones (Spencer and Davie, 2001;Zhang and Nelson, 1988). These are rapidly deacetylated when the inhibitors are removed. The second consists of mono- and di- acetylated histones which can be seen even in the absence of HDAC inhibitors and are deacetylated much more slowly and are termed class II acetylated histones.

It has been shown that the class I acetylated histones are associated with the transcriptionally active β-globin genes in chicken erythrocytes (Spencer and Davie, 2001). In addition, it is clear that the status of other histone modifications can influence acetylation. Many experiments have shown that serine 10 or 28 phosphorylated histone H3 are both subject to dynamic and continuous acetylation (Barratt et al., 1994;Clayton et al., 2000;Dunn and Davie, 2005;Dyson et al., 2005). Another study has shed light in the association seen between H3K4 methylation and histone acetylation in ChIP studies (Barski et al., 2007;Heintzman et al., 2007;Koch et al., 2007). It was found that H3K4 methylated histones were TSA hypersensitive, *ie* they belong to the class I group, but H3K9 methylated histones do not (Hazzalin and Mahadevan, 2005). Phosphorylation of histone H3 could be found on nucleosomes that were also K4 tri-methylated and acetylated.

Based on studies showing the half life of methyl groups on histones to be equivalent to those of the histones themselves, histone methylation was thought to be a relatively stable mark (Byvoet et al., 1972;Duerre and Lee, 1974). In addition, enzymes capable of demethylating histones were not identified until very recently. Even in early experiments, however, evidence was found for a small but detectable pool of methylation that was turning over (Borun et al., 1972). It was also shown that histone methylation was turned over very rapidly upon activation of the *pS2* gene promoter in human cells (Metivier et al., 2003).

The enzyme LSD1 was the first demethylase to be described. It has been found to demethylate both H3K4 and H3K9 (Shi et al., 2004;Metzger et al., 2005) but the enzyme does not appear to part of a widespread family of histone demethylases (Klose and Zhang, 2007). More recently, another enzyme with demethylase activity was purified, JHMD1A (Tsukada et al., 2006). This enzyme contains the Jumonji-C type domain which is widespread in eukaryotic genomes and a number of enzymes containing these domains have been identified as demethylases

specific for a variety of methylated histones (Klose and Zhang, 2007). Finally, demethylimination by the enzyme PADI4 has been shown to convert monomethyalted arginine residues to citrulline (Cuthbert et al., 2004;Wang et al., 2004).

1.3 Higher Order Chromatin Organisation

1.3.1 Euchromatin and Heterochromatin

In early studies of the nucleus, chromatin was divided into two types: euchromatin and heterochromatin. Heterochromatin was first defined as chromatin which did not decondense at the end of telophase, unlike the rest of chromatin which is termed euchromatin (Heitz, 1928). Even at this early stage it was proposed that heterochromatin might represent a functionally inactive portion of the genome. Heterochromatin can be defined as two sorts, constitutive heterochromatin which is constitutively condensed on both copies of a particular chromatin and facultative chromatin whose condensation state changes throughout development (Brown, 1966).

Constitutive heterochromatin is mainly associated with centromeric and juxtacentromeric regions which are made up of satellite tandem repeats. In electron micrographs of nuclei it can be seen as electron dense regions within the nucleus and when stained with 4,6-diamino-phenylindole (DAPI) it is seen as bright foci (especially in mouse cells). These regions stain more brightly with DAPI, not only because they correspond to dense regions of chromatin but because DAPI preferentially binds to AT-rich DNA sequences and satellite repeats are generally AT-rich.

The predominant satellite repeats in mouse cells are known as major and minor satellites. We know, from sedimentation analysis, that both major and minor satellite have a more condensed and ordered chromatin structure than bulk chromatin (Gilbert and Allan, 2001). In addition, immunofluorescence of nuclei with antibodies for the H3K9me3 modification show that it co-localises with DAPI bright

heterochromatic foci (Gilbert and Allan, 2001;Peters et al., 2003). The modification is also present at both major and minor satellite sequences by ChIP (Martens et al., 2005). GFP tagged linker histone H1 is also more abundant at these DAPI bright spots than in the rest of the nucleus (Misteli et al., 2000).

Another heterochromatin protein, HP1, was first described in *Drosophila* (James et al., 1989). It is now known that H3K9me3 marks deposited by the histone methyltransferase, Su(var)3-9, are responsible for recruiting this protein to heterochromatin (Lachner et al., 2001). In mouse and human, three different HP1 proteins are found, α, β and γ (Singh et al., 1991;Nielsen et al., 1999). All three localise to blocks of heterochromatin but HP1α is more tightly bound to chromatin and strongly localised to these sites (Gilbert et al., 2003;Remboutsika et al., 1999). In *Drosophila* artificially tethered HP1 can silence a transgene without corresponding recruitment of Su(var)3-9 and histone methylation (Li et al., 2003) but in a similar experiment in human cells found that repression corresponded to methylation of the targeted region (Verschure et al., 2005). When tethered in human cells, HP1 induced cytological compaction of a transgene array. By sedimenting nucleosomal arrays *in vitro,* it has been suggested that HP1α can compact arrays of unmodified arrays and associates better with arrays constructed from the histone variant H2A.Z (Fan et al., 2004). Another theory is that HP1 functions as a docking platform, recognising the H3K9me3 mark and recruiting other factors to heterochromatin (Li et al., 2002). Recently, four new heterochromatin proteins whose targeting is mediated by HP1 have been described (Greil et al., 2007) and HP1 has found to be associated with both transcriptionally active and inactive regions in flies (de Wit et al., 2007).

1.3.2 Genome Organisation in the Interphase Nucleus

Apart from blocks of heterochromatin, the other nuclear compartment easily identifiable with a microscope is the nucleolus. The rest of the interphase nucleus is not however, a random assortment of DNA and proteins. Within its space there are a

large number of different membrane free compartments (Spector, 2003). In addition, it is clear that chromatin is not randomly organised within the nucleus. Blocks of heterochromatin are clearly visible within the interphase nucleus and these are generally found around the nuclear periphery and surrounding the nucleolus (Sumner, 2003).

Replication of the genome is temporally regulated and certain genomic regions are known to replicate later than others. The latest replicating DNA sequences are those that are known to be present in blocks of constitutive heterochromatin. Synchronised populations of cells can be pulse labelled during S phase to determine the localisation of replicating DNA in the cell. This type of analysis shows that indeed late replicating heterochromatin is positioned around the periphery of both the nucleolus and nucleus (Ferreira et al., 1997). Similarly, early replicating DNA is found to localise to central regions of the nucleus. If early and late replicating DNA are differentially labelled, they are found to occur in a non-overlapping pattern, suggesting that they occupy different compartments within the nucleus (Sadoni et al., 1999;Zink et al., 1999). In general gene rich-regions of the genome replicate earlier than gene poor regions and the central compartment of the nucleus containing early replicating DNA also contains gene-rich DNA (Sadoni et al., 1999).

The localisation of histone modifications within the nucleus also shows a similar compartmentalisation. Acetylated histones are found in the interior, early replicating part of the nucleus and are excluded from the nuclear periphery (Sadoni et al., 1999;Gilchrist et al., 2004). Exclusion from the periphery is mediated by the action of HDACs as their inhibition abrogates this pattern (Gilchrist et al., 2004). It is not known however if this population of histones corresponds to the class I hyperacetylated histones discussed in *Section 1.2.4.4* (p17). As well as marking pericentromeric blocks of heterochromatin, H3K9me3 is also found in other nuclear regions in mouse, human and chicken (Gilbert et al., 2003). It is thought that these correspond to regions of facultative heterochromatin in the nucleus. H3K9me1, H3K9me2 and H3K27me3 are all diffusely distributed through the nucleus (Gilbert et al., 2003) and do not localise so strongly to heterochromatin. It is also known that different domains of H3K27 and H3K9 methylation on the inactive X chromosome

do not overlap in the interphase nucleus (Chadwick and Willard, 2004) and this organisation may extend to the nucleus in general (Zinner et al., 2006). The activating H3K4 methylation mark is also diffusely distributed and despite the coincidence of its ChIP profile with acetylated histones, does not appear to be excluded from the nuclear periphery (D. Sproul, Master's Dissertation, University of Edinburgh). However, all forms of H3K4 methylation are excluded from blocks of heterochromatin within the mouse nucleus.

1.3.3 Chromosome Territories

It has become clear that the chromosomes in the interphase nucleus are not arranged at random. Experiments using lasers to cause damage to chromosomes were the first to determine this. If a particular area of the nucleus is targeted by a laser, then damage is not spread between all chromosomes, but confined to a few (Zorn et al., 1979;Zorn et al., 1976). This suggests that each region of the nucleus is occupied by particular chromosomes rather than random DNA from all chromosomes. With the development of chromosome paints, the position of each chromosome could be visualised using Fluorescence *In Situ* Hybridisation (FISH) and it was shown that each chromosome occupies a discrete 'territory' within the nucleus (Manuelidis, 1985). The positions of these specific chromosome territories are conserved between different cell types and even between species (Croft et al., 1999;Boyle et al., 2001;Tanabe et al., 2002a;Tanabe et al., 2002b). In general, it was found that gene-rich chromosomes adopted positions that were, on average, more central than gene-poor chromosomes (Croft et al., 1999;Boyle et al., 2001). However, in some cells where the nuclei are flat and ellipsoidal, chromosome size appears to be a greater factor in determining their positioning, smaller chromosomes are found towards the centre and larger ones at the edge (Bolzer et al., 2005). Some chromosomes are preferentially found adjacent to each other, but this arrangement is probabilistic rather than absolute (Cornforth et al., 2002;Bolzer et al., 2005).

It was initially suggested that an interchromsomal space may exist between territories and that transcription occurred adjacent to this space on the surface of the territories (Cremer and Cremer, 2001;Zirbel et al., 1993). However, intermingling

has been seen between adjacent territories in the nucleus using high resolution techniques (Branco and Pombo, 2006) and DNA sequences from a particular chromosome are capable of adopting positions outside of the main body of its chromosome territory (Mahy et al., 2002a). In addition, studies of labelled chromatin components have shown that nearly all are highly dynamic within all parts of the nucleus (Bubulya and Spector, 2004) and labelled dextrans and macromolecules are capable of accessing even the interior of chromatin domains (Verschure et al., 2003). RNA synthesis also does not appear to be preferentially localised to the interchromatin space (Sadoni and Zink, 2004). Therefore, the arrangement of chromosomes into territories does not set up a system where steric hindrance prevents transcriptional machinery from accessing the interior of chromosome territories.

1.3.4 Organisation of Chromosome Territories

Many studies have looked at the position of individual loci relative to their parent chromosome territory by performing FISH with a genomic clone such as a BAC (~100-200Kb in size) and a chromosome paint. Examination of a region of human chromosome 11 suggested that its organisation was dictated by gene density rather than transcription and was conserved in the syntenic region of mouse cells (Mahy et al., 2002b;Mahy et al., 2002a). Gene-rich parts of the chromosome have a tendency to localise to the edge or outside their parent chromosome territories. These regions of the chromosome also appear to be enriched in 30nm fibres with an open conformation (Gilbert et al., 2004).

In *Drosophila*, FISH using probes for contiguous regions of the chromosome showed that gene poor and gene rich regions occupied non-overlapping regions in the nucleus (Boutanaev et al., 2005). A similar type of analysis has been conducted in mouse cells and similarly concludes that gene-rich and gene-poor regions of the genome occupy separate compartments in the nucleus (Shopland et al., 2006). In human cells, mixtures of BAC probes spanning ~5-10Mb portions of chromosomes have been used to measure the characteristics of gene-rich and gene-poor interphase chromatin structure (Goetze et al., 2007). Again, very little intermingling of gene-

rich and gene-poor regions was seen within the interphase nucleus. This is despite the authors concluding that the exact structure of each region was highly variable and dynamic. Gene-rich parts of the chromosome were also positioned closer to the centre of the nucleus on average, than gene-poor regions. These studies have all relied on the use of FISH to examine the nuclear positioning of particular genomic regions. By crosslinking nuclei using formaldehyde and isolating DNA sequences that are associated with each other, it is possible to assay the nuclear environment of a particular nuclear sequence using microarrays. Using this technique the authors of one recent study have come to similar conclusions as those utilising FISH, that gene-rich and gene-poor regions occupy distinct nuclear compartments (Simonis et al., 2006).

1.3.5 The Nature of Chromatin Organisation above the 30nm Fibre

When DNA is packaged into the 30nm fibre, this is sufficient to compact it ~40-50 fold (Sumner, 2003). However, within the nucleus higher levels of packing are found and so some mode of chromatin packing must operate above the level of the 30nm fibre. However, the nature of this structure is difficult to determine, mainly due to a paucity of techniques. Some electron micrograph studies of nuclei describes fibre like structures ~80 to 130nm thick (Belmont and Bruce, 1994). Pulse-labelling during replication has been used to provide evidence of chromatin existing in large scale domains of ~0.2-0.4μm in size, termed ~1Mb domains (Zink et al., 1998). Work on metaphase chromosomes has suggested a radial loop model in which 30nm chromatin fibres are arranged into loops by regular attachments to a scaffold like structure which forms the chromosomes backbone (Boy de la Tour and Laemmli, 1988).

One approach to model higher order chromatin structure is to hybridise pairs of genomic probes to the interphase nucleus (*Figure 1.2A*). The separation of these probes can then be measured in a large number of cell nuclei and the resulting distributions statistically analysed to model features of *in vivo* organisation. Using

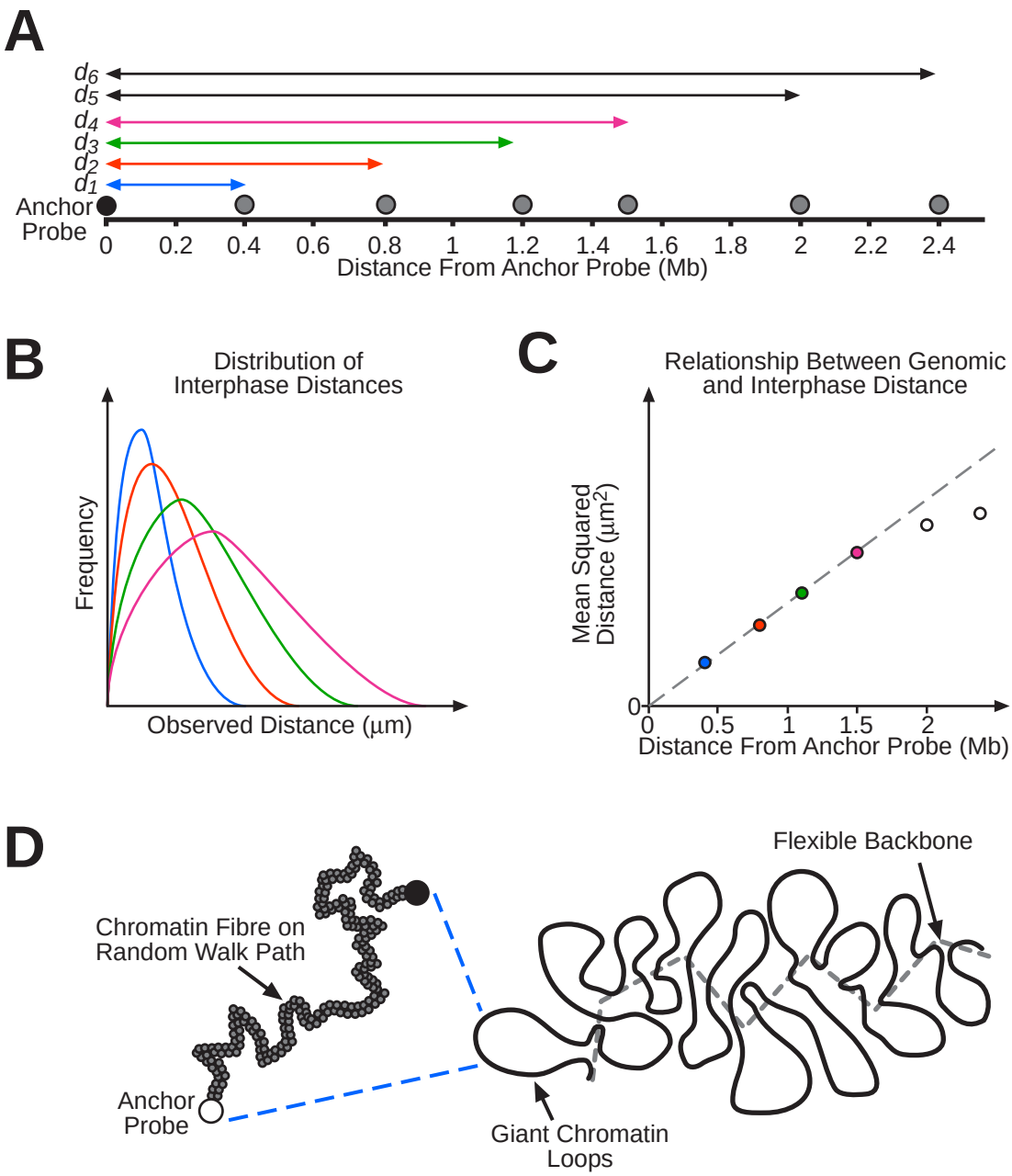


Figure 1.2 Measuring Interphase Chromatin Organisation Using a Random Walk Model

A. Schematic of a hypothetical region of a chromosome. Different pairs of genomic probes are used to measure interphase distance in the nucleus by FISH. Each has a common anchor probe and a second probe at a variable distance from this.

B. If FISH is performed with these probe pairs and the distance between signals measured in a large number of nuclei then the observed distances form a Rayleigh distribution for each probe pair (van den Engh *et al*., 1992).

C. For distances less than ~2Mb, there is a linear relationship between the Mean Square of the observed interphase distances and the genomic distance between probes. At distances greater than this the relationship breaks down (van den Engh *et al*., 1992). Data is hypothetical.

D. These observations suggested that chromatin fibres follow a random walk path through the nucleus. Based on observations over a large range of distances, it was also proposed that interphase chromosomes are composed of giant loops attached to a flexible backbone which follows a random walk path through the nucleus (Sachs *et al*., 1995, Yokota *et al*., 1995).

this approach it was found that over genomic distances between ~50Kb and ~2Mb the distributions of interphase distances suggested the chromatin followed a random walk path through the nucleus, behaving like a flexible polymer (*Figure 1.2B-D*, (van den Engh et al., 1992). By measuring interphase distances for a large range of genomic distances, it was suggested that chromatin was organised into a system of giant loops connected to a flexible backbone (*Figure 1.2D*, (Sachs et al., 1995;Yokota et al., 1995). Based on the relationships observed between genomic probes separated by larger genomic distances it was concluded that there are two levels of chromatin structure over the range of 0.15 to 200Mb (Sachs et al., 1995).

This model has been used to analyse different regions of the genome. It was possible to detect a difference in the compaction level with gene-poor regions being more compact than gene-rich ones (Yokota et al., 1997). In a recent study in our laboratory, these differences were also observed and found to correspond to biophysically measured differences in 30nm fibre structure (Gilbert et al., 2004). It is important to note that even regions found to be decondensed by this method do not decondense to the level of a 30nm fibre and they therefore measure a level of chromatin packing above that of the 30nm fibre.

1.4 Genomic Organisation

1.4.1 G and R Bands

A common theme running through analyses of higher order chromatin structure is that variation often corresponds to gene-rich and gene-poor regions. It is known that these regions are not distributed at random in genomes. By examining metaphase chromosomes using a variety of banding techniques, two general types of bands are observed. One set, known as R bands, contains gene-rich, early replicating DNA (Craig and Bickmore, 1994;Drouin et al., 1994). The reciprocal set of bands, G bands, are late replicating and contain fewer genes. G bands are named because they stain darkly with Giemsa after treatment with dilute trypsin. R bands are the reverse

set of bands which can be made to stain darkly using a different protocol (Sumner, 2003).

The complete sequencing of the human genome confirmed the gene-rich and gene-poor nature of R and G bands respectively (Lander et al., 2001). An R banding pattern can also be observed on metaphase chromosomes when immunofluorescence is performed using antibodies against acetylated Histone H4 (Jeppesen, 1997). This pattern was confirmed with the advent of genome-wide ChIP studies (Roh et al., 2005). A further larger study from the same group has extended the analysis to histone methylation, suggesting that H3K9me3 is found preferentially in G bands and H3K4me3 in R bands (Barski et al., 2007). However, H3K27me3 did not appear to so clearly demarcate chromosome bands despite also being associated with inactive genomic regions. Using microarrays, genome-wide studies of replication timing have also confirmed that in general, the gene-rich R bands replicate earlier than gene-poor regions (Woodfine et al., 2004). Finally, a study from our laboratory examined the distribution of biophysically purified 'open' and 'compact' chromatin fibres across the human genome (Gilbert et al., 2004). The presence of open chromatin fibres correlated with gene density and hybridisation of a DNA fraction enriched in open chromatin fibres to metaphase chromosomes revealed an T-banding pattern (T bands are a subset of R bands and are the most gene-rich regions of the genome).

1.4.2 Organisation of Genes within the Genome

Limited sequencing of genomes had revealed the existence of genomic regions containing clusters of genes that were specifically expressed in similar tissues. This led to the idea that the whole genome may be organised on such as basis (Sproul et al., 2005). However, analyses of complete genome sequences have suggested that this is not the case (Hurst et al., 2004). Whole genome expression analyses show that within the human genome, there are a number of genomic domains with a high average expression level called ridges (Versteeg et al., 2003;Caron et al., 2001). These are interspersed with gene domains of a similar size that have a low expression level, which have been termed antiridges. Genes in ridges tend to be more broadly

expressed than those in antiridges (ie expressed in more tissues and cell types) and it has been hypothesised that genome organisation is driven mainly by the clustering of broadly expressed house keeping genes into gene-rich regions (Lercher et al., 2002). One theory is that this arrangement may be driven by a need for essential house-keeping genes to be located in regions with low transcriptional noise (Batada and Hurst, 2007).

It is not known whether the differing chromatin structures associated with gene-rich and gene-poor regions drive the organisation of the genome or if the differential distribution of genes causes the variations of chromatin structure seen across the genome. However, a recent study found that ridges and antridges could influence gene expression (Gierman et al., 2007). Identical GFP transgenes were integrated as a single copy into random locations in the human genome and it was found that on average, those integrated into ridges had a higher expression level than those that had integrated into antiridges. It has also recently become apparent that a far larger portion of the genome than previously thought is transcribed but it is not currently known how total transcription levels relate to the patterns of ridges and antiridges seen for genic transcription (Birney et al., 2007;Kapranov et al., 2002).

1.4.3 Clusters of Co-ordinately Regulated Genes

Although the genome overall is based on the organisation of house-keeping genes into gene-rich and gene-poor areas, a number of clusters of structurally and functionally related genes also exist in the genome. The best known and most intensely studied are the Hox and Globin clusters, although others such as the Major Histocompatibility Complex (MHC), histone gene clusters and the olfactory gene clusters also exist. It is thought that they have a potential to exist in both gene-rich and gene-poor regions of the genome (Versteeg et al., 2003). Indeed in humans, the *β-globin* cluster is in a relatively gene-poor area whereas the *α-globin* cluster is in a gene-rich area (Brown et al., 2006b) and different *Hox* clusters have different genomic environments in the mouse (Morey et al., 2007). These clusters often

encode highly specialised genes that have strict temporal and spatial expression patterns. Despite their comparative rarity in the genome, much of what we know about how chromatin structure regulates gene expression comes from studies looking at gene clusters. This is because they make attractive model systems to study as chromatin structures can be compared between two cell populations, one where the cluster is active and another where it is inactive which allows to infer the characteristics of active and inactive regions of the genome in general.

1.4.3.1 The Globin Clusters

In the human genome there are two globin clusters: α-globin which is on human chromosome 16 and *β-globin* locus is on human chromosome 11. As mentioned above, they exist in very different environments but together they encode the protein products that make up haemoglobin. Therefore they have a very strict expression pattern that is confined to erythroid cells. It is thought that the *β-globin* arose due to an ancestral duplication of the *α-globin* locus (Czelusniak et al., 1982). In addition, the two loci were probably together at first and relics of this organisation can be found in marsupials (De Leo et al., 2005) and fish (Gillemans et al., 2003).

At both gene clusters there is a strict relationship between gene order and the order of temporal gene expression during development, with the first gene expressed in the embryo and later genes becoming active in the foetus and adult. The requirement of a strict order may exert selection pressure on the locus which maintains its organisation. In chicken however the *β-globin* locus has a different order, and the last gene is expressed first before expression resumes a more typical pattern. If the chicken locus is transfected into mice, the correct chicken temporal sequence is observed (Mason et al., 1995;Mason et al., 1996). In addition, the human and chicken loci probably arose independently from the duplication of a single ancestral gene and its organisation does not reflect a rearrangement of the mammalian locus (Czelusniak et al., 1982).

All the genes in the *β-globin* locus are transcribed in the same orientation (ie from the same DNA strand). The reason for this arrangement is unclear but antisense

transcripts extending through the locus have been observed to occur specifically in erythroid cells (Plant et al., 2001;Ashe et al., 1997). Another feature of the *β-globin* locus is an element known as the Locus Control Region (LCR). It consists of a number of DNase I hypersensitive sites (HSs) and is essential for efficient transcription of the genes. Using a method termed Chromatin Conformation Capture (3C), where adjacent loci in the nucleus are crosslinked with formaldehyde, it was shown that HSs of the LCR are in close spatial proximity to a further set of HSs lying close to the genes (Tolhuis et al., 2002). This has also been visualised and confirmed using other techniques (Carter et al., 2002). The two sets of HSs form the base of a looped domain to which globin genes are then sequentially recruited as they are expressed (Tolhuis et al., 2002). This chromatin architecture is erythroid specific and depends on erythroid transcription factors (Drissen et al., 2004;Vakoc et al., 2005). It has been found that in murine progenitor erythrocytes that are commited to globin expression but do not yet express the genes, some of the HSs and the LCR are associated but the globin genes themselves are not. As the genes become expressed they are recruited sequentially to the loop hub: the embryonic genes in primitive embryonic erythrocytes and the adult *β-globin* genes in erythroid cells from foetal liver (Palstra et al., 2003).

As one of the most intensely studied loci in the genome, we know a lot about the pattern of histone modifications at the *β-globin* locus throughout development. The first demonstration of a direct association of acetylated histone with active genes used the chicken *β-globin* locus as a model (Hebbes et al., 1988). As described in *Section 1.2.4.4* (p17), many of the histones at the locus are subject to the constant action of histone acetyltransferases and HDACS and belong to the class I group of acetylated histones. The action of HDACs appears to affect the antisense transcription through the locus as TSA can be used to induce antisense transcription through the locus, even in cell types which have an inactive locus (Plant et al., 2001). The *β-globin* LCR has been shown to hyperacetylated during many developmental stages (Forsberg et al., 2000) but the genes themselves are differentially acetylated throughout development. In the embryonic yolk sac, acetylated histones are found at both the active embryonic genes and the inactive adult genes. Later, in the foetal liver, acetylated histones are retained at the promoters of adult genes and are lost

from the inactive embryonic genes. Methylation of Histone H4 arginine 3 is a necessary initial step for the subsequent acetylation of histones H3 and H4 (Huang et al., 2005). Tagged Histone H3.3 is also incorporated at active gene promoters in the locus, but more strongly at the upstream hypersensitive sites and LCR (Jin and Felsenfeld, 2006). Histone modifications have also been studied at the *α-globin* locus (de Gobbi et al., 2007). Housekeeping genes surrounding the locus are constitutively marked by H3K4me and acetylation whereas promoters and regulatory elements within the locus itself are only marked with these activating modifications in erythroid cells.

The interphase organisation of the globin clusters has also been studied and their nuclear positions were found to depend on genomic context rather than on their transcriptional status (Brown et al., 2006b). The human *α-globin* locus is generally found outside of its chromosome territory whereas the human *β-globin* locus and the mouse *α-* and *β-globin* loci are found inside their parent chromosome territory, irrespective of their transcriptional activity (Brown et al., 2006b;Mahy et al., 2002a). However, the human *β-globin* locus was found to associate with centromeric α-satellite heterochromatin when inactive in primary B and T cells but not when active (Brown et al., 2001). In contrast, the *α-globin* was not associated with heterochromatin, even when inactive. A recent study making use of 3C technology coupled with microarrays (4C) also suggested that the active *β-globin* locus was generally associated with active loci in the genome and the inactive locus associated with other transcriptionally inactive regions in the genome (Simonis et al., 2006). These were mostly regions from the same chromosome but included some regions from other chromosomes.

In addition, the *β-globin* locus has been analysed at the level of its secondary chromatin structure, *ie* 30nm fibre conformation. In a biophysical analysis of chromatin structure, it was shown that a restriction fragment corresponding to the *β-globin* locus moves more slowly through sucrose gradients than bulk chromatin of a similar size (Kimura et al., 1983). Sucrose gradients separate particles based on two parameters, their mass and their shape. The behaviour of the *β-globin* locus suggested that it adopted an unfolded structure and was specifically seen in chicken erythrocytes (Kimura et al., 1983) and in an independent study (Fisher and

Felsenfeld, 1986). When this behaviour was further characterised, it was found that it could be explained by the loss of a single nucleosome at the HSs which created discontinuities in the chromatin fibre (Caplan et al., 1987).

1.4.3.2 Hox Clusters

Another set of well studied gene clusters in the genome are the Hox gene clusters. In vertebrates there are 4 *Hox* clusters, known as *HoxA, B, C* and, *D* (Duboule, 2007). All of the clusters are thought to have originated from a single ancestor but the the number of clusters present in the genome and the number of genes within each cluster varies between organisims (Simonis et al., 2006;Ferrier and Minguillon, 2003;Hoegg and Meyer, 2005). Hox genes are regulated in both a temporal and spatial manner. It has been shown that correct spatial expression patterns are maintained for *Hox* transgenes inserted at ectopic locations in the genome but that, temporal regulation is lost (Kmita and Duboule, 2003). In addition, studies of the organisation of *Hox* genes in the urochordate *Ciona intestinalis* have shown them to be partially dispersed throughout the genome (Ikuta et al., 2004). Spatial control of gene expression still occurs along the larval neural tube but temporal co-ordination appears to be lost. In *Oikopleura dioica* the genes are completely dispersed but, anterior-posterior patterning of *Hox* expression still appears to occur (Patel, 2004;Seo et al., 2004). This might suggest that the clustering of the *Hox* genes is important for their temporal regulation but not their spatial regulation.

The genes at *Hox* loci are all transcribed in the same orientation. Many non-coding RNAs have also been found to be produced from the intergenic regions of the *Hox* loci and, like those at the *β-globin* locus, they are generally transcribed in an antisense fashion to the genes (Mainguy et al., 2007;Sessa et al., 2007). It was proposed that this transcription served to open the chromatin state of the cluster and remove repressive polycomb proteins and the H3K27 methyl marks, thereby allowing the activation of adjacent genes (Sessa et al., 2007). A further study made use of high resolution microarrays, characterised a non-coding RNA from the *HoxC* cluster termed HOTAIR (Rinn et al., 2007). Rather than regulating the *HoxC* cluster,

it was found that this RNA mediates a repressive effect of part of the *HoxD* cluster via the polycomb repressive complex.

Hox clusters do not generally appear to be regulated by LCR-like regulatory elements. However there is a Global Control Region (GCR), located 5' of the *HoxD* cluster that is believed to regulate expression of the 5' *HoxD* genes in the distal limb bud (Spitz et al., 2003). Its mode of action is unknown but it has been found to have an effect on genes that are unrelated but adjacent to the cluster. Recently, it has been proposed that anterior-posterior collinear expression of the same cluster is controlled by a further element an early limb control region (ELCR) lying 3' of the cluster (Zakany et al., 2004). The GCR probably represents a relatively recent evolutionary event at the *HoxD* cluster because no similar elements are known to exist at other Hox clusters (Spitz et al., 2003).

Initial analyses of the histone modification patterns at *Hox* genes suggested they followed similar patterns to many other genes, with acetylation marks present at their promoters when active (Chambeyron and Bickmore, 2004). However, more recent large-scale studies of histone modifications have described large scale domains of histone H3K4 methylation stretching across multiple active genes at the *Hox* loci, in clear contrast to the punctate localisation of this modification to active promoters in other parts of the genome (Bernstein et al., 2005). Another study found that at human *Hox* loci, inactive portions were marked by broad domains of H3K27 trimethylation and bound by polycomb proteins whereas active portions were defined by broad domains of transcription and presumably H3K4 methylation (Rinn et al., 2007).

The nuclear organisation of *Hox* loci has also been studied. Firstly it was shown that in an ES cell based system the mouse *HoxB* locus moves out of its chromosome territory upon activation of gene expression (Chambeyron and Bickmore, 2004). Using a pair of genomic probes, it was also shown that this is accompanied by a remarkable chromatin decondensation. Higher order folding was not apparent and the distance separating the probes corresponded to that expected from a 30nm chromatin fibre. Similar reorganisations have been seen in embryos (Chambeyron et al., 2005) and at the *HoxD* locus which is known to have a different genomic organisation (Morey et al., 2007). However, the repositioning and

decondensation of the of the *HoxD* locus were shown to be distinct and not merely different manifestations of the same process. A study using the 4C technique also found that upon gene activation, the *HoxB* locus was more frequently found associated with sequences that were not part of its chromosome, confirming that upon activation it loops out from its parent chromosome territory (Wurtele and Chartrand, 2006).

1.4.3.3 Superclusters

Clusters of co-regulated genes tend to be ~50-200Kb in size. Another type of cluster exists in the genome which are much larger and have been termed superclusters (Sproul et al., 2005). These are large areas of the genome which contain clusters of gene families, some of which are functionally related. The prototypical supercluster is the Major Histocompatibility Complex (MHC) which stretches across 7.6 Mb of human chromosome 6 and is the most gene-dense region of the genome (Horton et al., 2004;Lander et al., 2001). It was first defined in 1936 and intensely studied due to its role in the immune system which has led to the proposal that the need to maintain haplotypes in linkage disequilibrium has kept genes clustered in the region (Trowsdale, 2002). However, the original definition of the MHC has been extended to include clusters of olfactory receptor (OR), tRNA and histone-coding genes that are not obviously functionally linked to immune system function (Horton et al., 2004). In general most of the region has a high transcriptional activity but the OR genes are silent in most cell types and even when active only one allele of one gene from the entire OR family is expressed in a particular cell (Shykind, 2005). In teleost fish the class I and class II MHC are separated in the genome, arguing that the region can still function without linkage (Kumanovics et al., 2003).

In contrast to the *Hox* and *globin* loci, little is known about the regulation of gene expression at the MHC. It is known that HSs exist ~9Kb upstream of the MHC class II Ea (HLA-DRA) gene but these appear to regulate that particular gene rather than the locus as a whole (Carson and Wiles, 1993). Regulatory motifs termed S-Y elements are also found close to the promoters of each of the class II genes (Krawczyk et al., 2004). These bind the regulatory factor X complex (RFX) and

class II transactivator (CIITA) which is induced by interferon Gamma. They are also present at the class II-associated invariant chain gene, Ii, located on a different chromosome (Krawczyk et al., 2004). CIITA associates with histone acetyltransferases (HAT) and has intrinsic HAT activity which results in domains of Acetylation around active class II genes (Masternak et al., 2003;Raval et al., 2001). These domains are, however, smaller and distinct from the domains observed at *Hox* loci.

It is known that the classical MHC is located outside of the chromosome 6 territory when active in lymphoblastoid cells (Volpi et al., 2000). This organisation is not seen for the extended MHC, or in cells where the MHC is inactive but can be induced by Interferon-γ. More recently, it was shown that the Interferon-γ induced looping is mediated through the transcriptional activator P-STAT1 and the chromatin remodelling factor BRG1 (Christova et al., 2007). Looping can also be induced by treatment with the HDAC inhibitor sodium butyrate and is accompanied by a cytological decondensation of the classical MHC (~3.4Mb in size). It was previously found that treatment with another HDAC inhibitor, Trichostatin A (TSA), did not induce the *HoxB* locus to loop out of its chromosome territory when inactive (Chambeyron and Bickmore, 2004) so their may be mechanistic differences between the looping at the MHC and Hox clusters.

1.5 The Epidermal Differentiation Complex

First defined in 1996, the Epidermal Differentiation Complex (EDC) spans roughly 2Mb of human chromosome 1q21.3 and at present has ~65 genes assigned to it (based on RefSeq Annotation, NCBI 36, (Mischke et al., 1996). These genes are divided into three groups based on structural similarity and are nearly all expressed during epidermal differentiation. The first group consists of 11 Small Proline Rich (*SPRR*) genes (Cabral et al., 2001), 16 Late Cornified Envelope (*LCE*) genes (Marshall et al., 2001), *Loricrin, Involucrin* and, the *NICE-1* gene (Zhao and Elder, 1997). The second group consists of 16 genes of the S100 family which are small calcium-binding, regulatory proteins whose genes flank the EDC (Marenholz et al., 2004). Finally the 'fused' genes which combine structural features of the other 2

groups and includes *Profilaggrin, Trichohyalin, Repetin (Krieg et al., 1997), Hornerin* (Makino et al., 2001) and *Cornulin* (Contzler et al., 2005). Most of these genes are arranged into clusters by type (*Figure 1.3*). Three other genes that do not appear to have direct roles in the epidermis (*PGLYRP3, PGLYRP4* and, *MCSP*) are also present in the region (Aho et al., 1996;Mathur et al., 2004).

By examining the region in genome browsers such as Ensembl or the University of California Santa Cruz (UCSC), the EDC appears to be less conserved than its surrounding genomic regions (*Figure 1.4*). However, conservation in individual mammalian species is higher than in non-mammalian species. A relatively low overall level of conservation is seen in the Oppossum and Platypus but closer examination of the sequence reveals that blocks of conservation correspond to most genes (data not shown).

We know that the human and mouse loci both have very similar organisation but there are differences in the number and order of *SPRR* genes that have been attributed to the fact that the genes have undergone independent expansion in the mouse and human lineages (Patel et al., 2003). Curiously, there is a report that the EDC may have originated in a genomic area that represents an ancient duplication of the MHC (Shiina et al., 2001). The MHC itself is believed to have evolved from an ancestral region predating chordates and appears to have been duplicated multiple times during mammalian evolution (Castro et al., 2004;Danchin et al., 2004). Although the EDC itself does not have a homologue on chromosome 6 the regions flanking it contain a histone cluster, tRNA genes and olfactory receptors like the MHC (Shiina et al., 2001).

Antibodies against loricrin and filaggrin detect a protein of the correct size in all amniotes (mammals, lizards and birds, (Alibardi and Toni, 2004). However, in amphibians there is only tentative evidence of a filaggrin like protein and neither filaggrin nor loricrin have been thus far found in fish (Alibardi, 2002;Alibardi et al., 2003). Sequences showing homology to SPRR proteins have also been found in the chicken genome (Martin et al., 2004). S100A proteins have been described in a number of mammalian species but are absent from invertebrates (Marenholz et al., 2004). Therefore, current evidence would suggest that some non-mammalian species

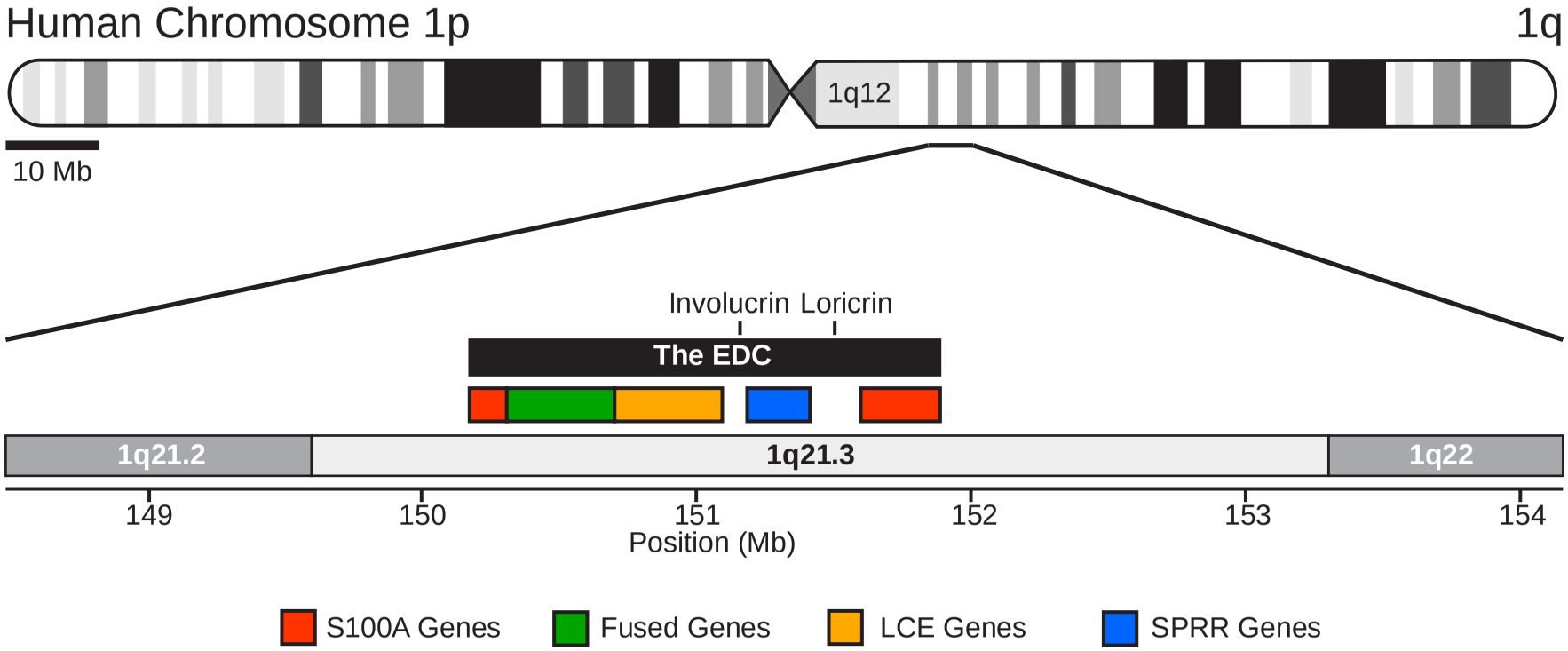


Figure 1.3 The Genomic Location and Structure of the Human Epidermal Differentiation Complex

An ideogram of human chromosome 1 with the location of the EDC indicated. The region is magnified and the locations of the gene clusters indicated by coloured blocks. Additional genes that are not part of the gene clusters are also illustrated. The genomic positions are taken from NCBI Build 36.

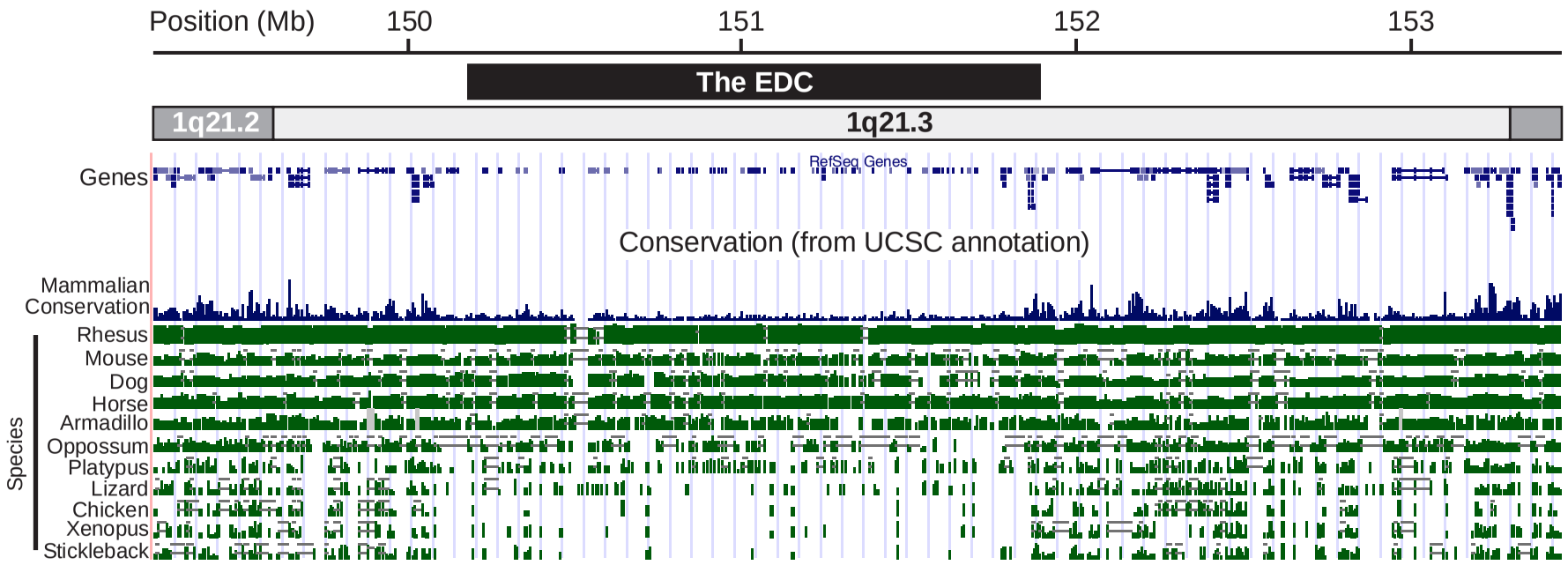


Figure 1.4 Evolutionary Conservation of the EDC

Map of the EDC showing evolutionary conservation as annotated in the UCSC browser. An ideogram of the region showing the chromosomal bands and the position of the EDC is drawn above the UCSC annotation. Genes are indicated below this as blue boxes (taken from the Refseq annotation). A graph of conservation in mammals is shown in blue, followed by separate graphs for a number of different species. These conservation graphs all come from the the UCSC conservation track. Genomic positions are from NCBI 36 and the UCSC browser build from March 2006 was used.

may posses the genes found in the human EDC but that they may not be organised into a cluster.

1.5.1 The Cornified Envelope

The basal layers of the epidermis are populated by asymmetrically dividing cells epidermal stem cells (Lechler and Fuchs, 2005) which produce a population of differentiating cells, keratinocytes, that move outwards through the granular layer and undergo a unique form of cell death, becoming the enucleated, flattened cells of the outer cornified layer (*Figure 1.5B*, (Candi et al., 2005). During this process an insoluble protein-lipid polymer structure, known as the cornified envelope, replaces the cell's plasma membrane (*Figure 1.5C*). The cornified cell layer acts as the body's barrier, excluding foreign substances and preventing the loss of vital fluids.

There are two major structural protein families expressed in keratinocytes (Mischke, 1998). The first of these are the keratins of which are part of the cellular cytoskeleton and are assembled into 10nm keratin intermediate fibres. The fibres are aggregated in differentiated cells by the action the protein filaggrin, collapsing the cell (Candi et al., 2005). The cornified envelope is formed when the second class of structural proteins (including: involucrin, loricrin, trichohyalin and the SPRR proteins) are crosslinked to each other and to the keratin bundles (*Figure 1.5B*, (Candi et al., 2005). A complex series of lipids are also synthesised and become crosslinked to the proteins by transglutimases (TGs). These enzymes are essential and *TG1* -/- mice die at birth due to a defect in barrier function (Matsuki et al., 1998). In contrast, when structural components of the cornified envelope are knocked out relatively mild phenotypes are observed, probably due to some redundancy in their functions and the upregulation of other proteins to compensate for the loss of one (Djian et al., 2000;Koch et al., 2000).

The main protein components of the cornified envelope are Involucrin, Loricrin and the SPRRs (*Figure 1.5B*), all of which are encoded in the EDC. Involucrin has been proposed to act as a scaffolding protein because it is an early component of the envelope and it is promiscuously cross-linked to the other structural proteins (Rice and Green, 1977; Yaffe et al., 1992). Some of the S100A

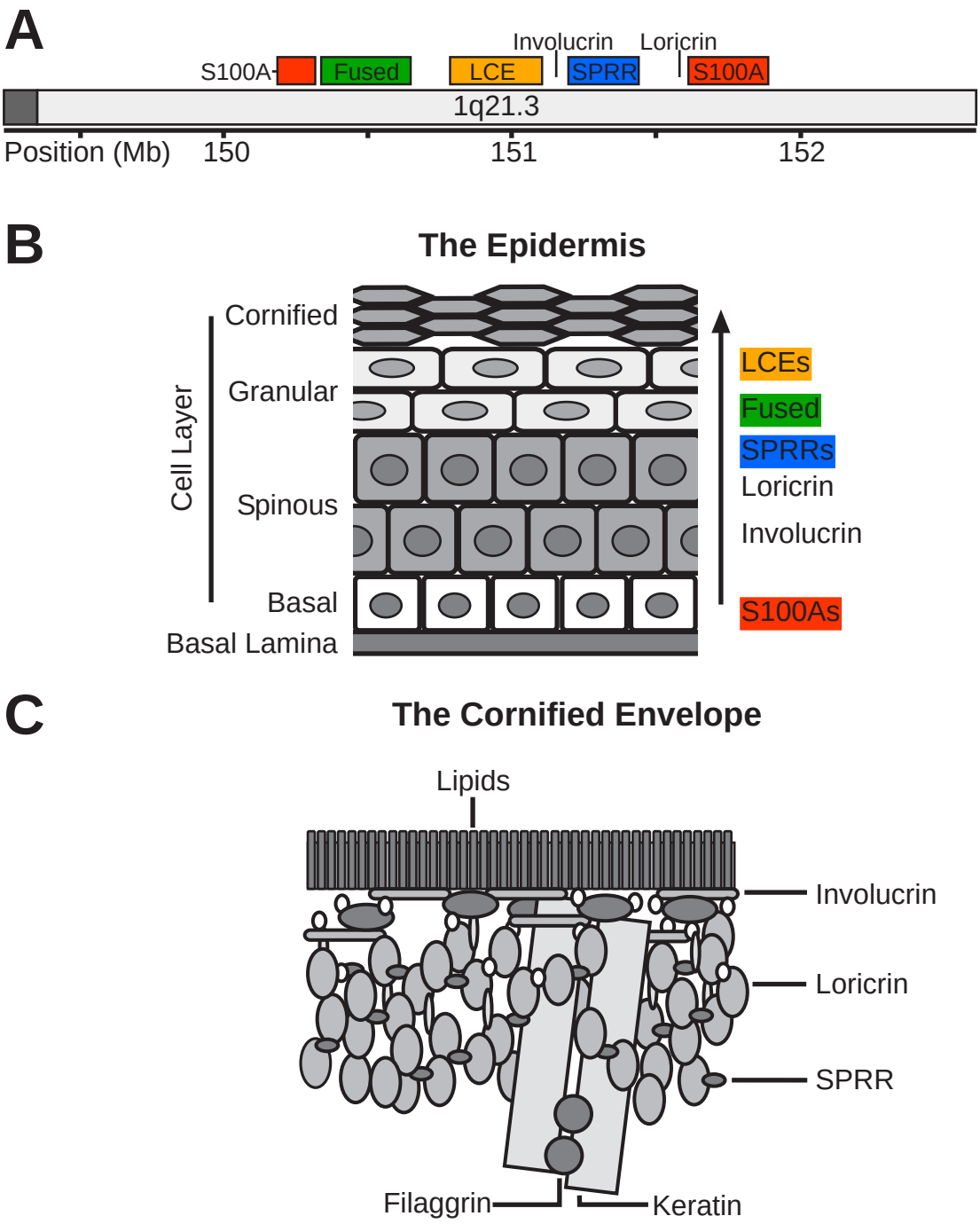


Figure 1.5 The Epidermis, cornified Envelope and Gene Expression at the EDC

A. The genomic structure of the gene clusters of the EDC, the ideogram is a simplified version of *Figure 1.1*. The different gene clusters are colour coded to indicate their expression in panel B. Genomic positions are from NCBI build 36.

B. The histology of human epidermis. Four distinct cell layers are depicted and the rough timing of gene expression for genes in the EDC indicated. Adapted from Mischke, 1998. The arrow indicates the direction differentiating cells move in, with the upermost layer being fully differentiated.

C. The structure of the cornified envelope. The major protein and lipid components of the cornified envelope are depicted and named, this structure replaces the cellular plasma membrane during terminal epidermal differentiation. Adapted from Candi *et al.*, 2005.

proteins have also been found to be crosslinked into the cornified envelope (Robinson et al., 1997). The SPRRs are a family of small proteins (6-18KDa) that can be divided into 3 classes. They have a common structure consisting of a head and tail domain separated by a central domain composed of 2-20 repeating peptide units (Cabral et al., 2001). Both loricrin and the SPRR proteins are predicted to have a relatively disorganised structure that confers elasticity to the cornified envelope (Candi et al., 1999;Candi et al., 1995). In the cornified envelope the ratio of loricrin to SPRR proteins varies from between around 1:100 to 1:3. Differing ratios of the two protein types appears to alter the physical properties of the epidermis in a manner reminiscent of composite materials (Steinert et al., 1998a;Steinert et al., 1998b). Different SPRR proteins may also have different physical properties allowing the epidermal tissues to be precisely adapted to their environment (Candi et al., 2005). The hypothesis that the cornified envelope is a structure formed to provide waterproof skin as an adaptation to living on land may explain why the EDC does not appear to be present in fish and amphibians (Alibardi, 2003).

1.5.2 The Epidermal Differentiation Complex and Disease

Abnormal function of the EDC is thought to be involved in a number of different disease states. Firstly a number of the members of the S100A clusters have been seen to be misexpressed in cancers. For example, gastric tumours overexpress many *S100A* genes (El-Rifai et al., 2002), *S100A6* and *A8* have been seen to be overexpressed in pancreatic cancers (Logsdon et al., 2003;Shen et al., 2004), *S100A2* was seen to be overexpressed in lymphomas (Hsieh et al., 2003) and in non-small cell lung tumours (Heighway et al., 2002). It has been hypothesised that these changes may be caused by the frequent rearrangements and deletions in the chromosomal region of the EDC that are observed in tumours (Marenholz et al., 2004). Unlike other genes in the EDC, the *S100A* genes are expressed in a range of cell types and are involved in a number of cellular processes including contraction, motility, cell growth, differentiation, cell cycle progression, transcription and

secretion (Marenholz et al., 2004). They function by binding to target proteins, and often require first to be activated by a calcium induced conformational change (Bhattacharya et al., 2004).

The reason why S100A proteins should be encoded in the EDC is slightly mysterious. There have been a number of studies that have examined their localisation in the epidermis and it seems that most of the *S100A* genes are expressed in the normal epidermis (Broome et al., 2003;Boni et al., 1997;Benoit et al., 2006). It is clear that the S100 proteins are able to act as calcium sensors in cells and it is known that intracellular calcium levels play an important role in the epidermal differentiation pathway (Hennings et al., 1980;Menon et al., 1992). S100A11 has been found in to inhibit growth of cultured human keratinocytes via $p21^{CIP1/WAF1}$ upon exposure to high calcium levels (Sakaguchi et al., 2003) and S100A8 has been reported to be an inhibitor of telomerase activity in a calcium dependent manner in HaCaT cells (Rosenberger et al., 2007). However, the proteins also play other roles in epidermal function: S100A10 is insensitive to calcium because of mutations in its calcium binding domains (Gerke and Weber, 1985;Saris et al., 1987) and S100A10 and A11 have been found to be incorporated into the cornified envelope (Robinson et al., 1997). S100A8/9 heterodimers are secreted from keratinocytes and may play chemo-tactic role in inflammation (Ryckman et al., 2003;Thorey et al., 2001).

Ultraviolet (UV) irradiation of the skin also causes transcriptional changes in the EDC and both the *SPRR* and *LCE* genes are generally upregulated in response to UV (Cabral et al., 2001;Jackson et al., 2005) along with the *S100A8* gene (Grimbaldeston et al., 2003). It is known that the *SPRR1* is regulated by interleukins 1 and 3, which are connected with UV irradiation (Yaar et al., 1995).

As would be expected from a locus which primarily functions in epidermal differentiation, a number of skin diseases have been linked to the EDC. Skin disease accounts for a very large proportion of disease and its frequency is probably higher than disease of any other organ (Rees, 2007). A defect in the synthesis of the filaggrin protein was observed in patients with the skin disorder ichthyosis vulgaris (Online Mendilian Inheritance in Man (OMIM) Number:146700, (Sybert et al., 1985). This is a common disease which, based on one survey, may have an incidence of 1:250 in English schoolchildren (Wells and Kerr, 1966). The disease was linked

to the EDC in a mapping study (Compton et al., 2002) and more recently two null mutations were found to be present in a number of patients with severe ichthyosis vulgaris (Smith et al., 2006). Surprisingly these two mutations were found in ~10% of people of European origin and were also linked to two other diseases Atopic Dermatitis or Eczema (OMIM: 603165) and Eczema-associated asthma (Palmer et al., 2006). Penetrance was not complete for either of these conditions and was lower for eczema-associated asthma than for atopic dermatitis. Atopic dermatitis is a very common disorder that may have a cumulative incidence of ~30% in the population (Rees, 2007). More recently, further null alleles have been uncovered and are also associated with these conditions (Nomura et al., 2007;Sandilands et al., 2007).

Psoriasis (OMIM 177900) is another disease of the skin which affects ~2-3% of the population (Rees, 2007). It is a complex trait that has been associated with a number of different loci, however, on of these is chromosome band 1q21, the site of the EDC (Bowcock and Cookson, 2004). The filaggrin null alleles associated with ichthyosis vulgaris and atopic dermatitis are not associated with psoriasis (Zhao et al., 2007). Although the gene responsible for this linkage has not yet been identified, it is clear that genes present in the EDC play a role in the pathogenesis of the disease. Studies of the expression and localisation of the S100A proteins have found much higher levels of S100A7, A8 and A9 in psoriasotic skin than normal skin (Broome et al., 2003;Benoit et al., 2006). The expression of the SPRR 1 and 3 proteins is also altered in a number of skin diseases including ichthyosis vulgaris and psoriasis (Koizumi et al., 1996).

1.5.3 Gene regulation in the Epidermal Differentiation Complex

Since the organisation of the EDC is conserved between mouse and man, it was proposed that the cluster is arranged to facilitate co-ordinate gene regulation (Mischke, 1998;Zhao and Elder, 1997) and *Figure 1.5A*). There have been many studies examining the role of different factors in regulating epidermal differentiation (for reviews see (Fuchs and Raghavan, 2002;Dai and Segre, 2004) and further

studies have examined the regulation of individual genes through a number of transcriptional factors, for examples see (Jang and Steinert, 2002;Presland et al., 2001;Jang and Steinert, 2002). However, few studies have addressed the overall regulation of the region.

Elements similar to a characterised binding site for the transcription factor *Wnt-5A* have been found in the EDC and proposed as putative regulatory elements (Morgan et al., 1999;Morgan, 2002). It was therefore hypothesised that this transcription factor may regulate the region but to date there have been no studies linking the transcription factor to the region or confirming that these are indeed functional *Wnt-5A* elements. Another, transcription factor that has been implicated in the control of many genes in the EDC is the Autoimmune regulator (AIRE, (Mathis and Benoist, 2007). It has been proposed that genes controlled by AIRE are clustered at various chromosomal locations including the EDC (Johnnidis et al., 2005). However, one problem with this study is that despite initial reports suggesting a broad pattern of expression for AIRE, it is now generally agreed that it is only expressed in lymphoid organs, particularly the thymus (Mathis and Benoist, 2007). This is obviously a rather different pattern from the genes of the EDC, therefore, the misregulation of these genes in the AIRE knockout mice may be due to secondary effects. Since it is known that many genes in the EDC are regulated as part of an immune or inflammatory response, perhaps in the knockout mouse the general defects in the immune system somehow lead to an immune response which causes altered expression in the EDC. Another potential regulator of the EDC might be the AP-2ε, a member of the AP-2 family of transcription factors (Tummala et al., 2003). However, so far, all that is known about this protein is that it is very specifically expressed in skin. However, the AP-2 family are thought to play some role in the regulation of the *Loricin* gene (Jang and Steinert, 2002).

Even fewer studies have examined the chromatin structure of the EDC. It has been shown that treatment with sodium butyrate (an HDAC inhibitor) or 5-azacytidine (which broadly causes DNA demethylation) leads to the upregulation of some *S100A* genes, the *SPRR* genes, *loricrin* and *involucrin* (Elder and Zhao, 2002). Recently, the promoter of *S100A6* was shown to be marked by H3K9me3 and H3K27me3 when inactive and H3ac when active in cultured cells (Lesniak et al.,

2007). Both the *SPRR* and *LCE* clusters can be upregulated by UV irradiation, implying they may be governed by global regulators (Cabral et al., 2001;Jackson et al., 2005). Patel *et al*., described the upregulation of the *SPPR* cluster in a mouse genetic model in which the transcription factor Klf4 was disrupted (Patel et al., 2003). In a further study, they identified two clusters of conserved HSs one of which showed enhancer like activity under *in vitro* differentiation conditions that upregulate the entire cluster of *SPRR* genes (Martin et al., 2004). This up-regulation does not affect the adjacent *Involucrin* and *Loricrin* genes. Finally, it has been shown that the EDC loops away from the chromosome 1 territory in human primary keratinocyte cultures, and that this does not occur in lymphocytes (Williams et al., 2002).

1.6 Aims of This Thesis

Through the ENCODE project and other large scale studies we have begun to understand how histone modification patterns correlate to transcription in the genome. These studies have shown that active gene promoters and regulatory elements are marked with activating modifications and when inactive, they are marked by repressive modifications. In general, large scale domains of modified histones do not occur except at some specific loci (Bernstein et al., 2005). At larger scales, studies of secondary and tertiary chromatin structure reveal that the human genome as a whole is organised into large scale gene-rich and gene-poor chromatin domains. This organisation corresponds to the organisation of the human genome into G and R bands.

However, we do not yet understand how these two distinct levels of organisation relate to each other. Analysis of data from the ENCODE project has found evidence for the existence of chromatin domains ~100Kb in size using ChIP, expression and replication timing data (Birney et al., 2007;Thurman et al., 2007). However, we do not know if such an organisation of primary structure corresponds to the organisation of secondary chromatin structure. The only study of chromatin fibre structure across the genome was conducted at too low a resolution to determine if similar domains exist (Gilbert et al., 2004). The aim of this project is to analyse the chromatin structure of a large genomic region to deduce the relationship between

the three levels of chromatin structure and to examine how the local pattern of histone modifications relates to the organisation of higher order chromatin domains.

To achieve this I could feasibly examine almost any region of the genome. However by examining a differentially regulated locus I can also compare the inactive and active locus to study the effect of changes in transcription in medium and large scale chromatin organisation. Classical gene clusters such as the *Hox* or *β-globin* loci span ~100Kb and are probably too small for this type of analysis. However, superclusters stretch across much larger genomic distances and so I have elected to study the EDC which is ~2Mb in size.

To comprehensively analyse chromatin structure across a region of this size I chose to combine biochemical techniques such as ChIP and sucrose gradient sedimentation with genomic microarrays. In collaboration with a commercial company, I have designed a tiling oligonucleotide array which covers the EDC to a resolution of ~1000bp. Large scale biochemical analyses preclude the use of tissues or primary cells because it is not easy to have a sufficiently large cell population from which to prepare chromatin for analysis. Cell lines, while obviously more artificial, allow large amounts of material to be isolated from a pure population of cells.

Therefore, the first stage in the study characterises the expression status of the EDC in a number of candidate cell lines to determine which are suitable models for further study (*Chapter 3*). Ideally, I wanted to find a keratinocyte cell line expressing many genes in the EDC and a control cell line with an inactive locus. Having selected my model cell lines I then characterised higher level chromatin organisation using FISH (*Chapter 4*) before examining primary and secondary chromatin structure using ChIP and sucrose gradient sedimentation combined with microarrays (*Chapter 5*). I then attempted to relate the different levels of chromatin structure to each other and examine whether the region is organised into one large chromatin domain or built up of many smaller ones. Furthermore, the EDC is obviously an important locus for the process of epidermal differentiation and its aberrant function plays a role in human diseases. It has been proposed that there is some mechanism of coordinate control at the locus (Mischke, 1998). However, few studies have investigated this possibility and by studying chromatin structure at the

locus, albeit *in vitro*, I hoped to gain some insight into whether such mechanisms exist and what they might be.

Although the EDC itself is rarely studied, there have recently been a number of large studies which have examined gene expression and chromatin structure across the human genome. 30nm chromatin fibre structure was analysed across the entire lymphoblastoid cell genome (Gilbert et al., 2004) and genome wide expression data is also available for lymphoblastoid cells (Goetze et al., 2007). The single study of the interphase chromatin organisation at the EDC used lymphoblastoid cells as a model of the inactive locus (Williams et al., 2002). A recent study has also examined the genome wide distribution of histone modifications in $CD4^+$ T-cells where we might expect chromatin structure across the genome to be similar to that found in lymphoblastoid cells (Barski et al., 2007). Therefore we can build up a picture of the inactive EDC using this data (*Figure 1.6A*). This would suggest that the inactive EDC has a low median gene expression levels and is depleted in H3K4me3 but enriched in H3K9me3 marks. This is in contrast to the regions flanking the locus which have the opposite pattern of histone modifications and are more active. In addition, the EDC and its proximal flanking region may have a compact secondary chromatin structure than the distal flank which is more open (*Figure 1.6A*, (Gilbert et al., 2004). However this data is at a low resolution and the status of the EDC and proximal flanking region are based on the values of just two microarray probes. From these data and what we know of nuclear organisation in the region in lymphoblastoid cells, I would suggest that the inactive EDC is marked by repressive histone modifications, has a condensed chromatin structure and is positioned inside the chromosome 1 territory (*Figure 1.6B*). We might expect that upon activation, the repressive marks will be replaced by activating marks, the chromatin will decondense and the locus will loop out of the territory. The flanking regions on the other hand will likely have a relatively decondensed structure in both cell lines and positioned on the edge or outside of the chromatin territory and be associated with active histone modifications. These data may also suggest that the EDC is represented by a single large chromatin domain with G band like properties when inactive but with R band like properties when active.

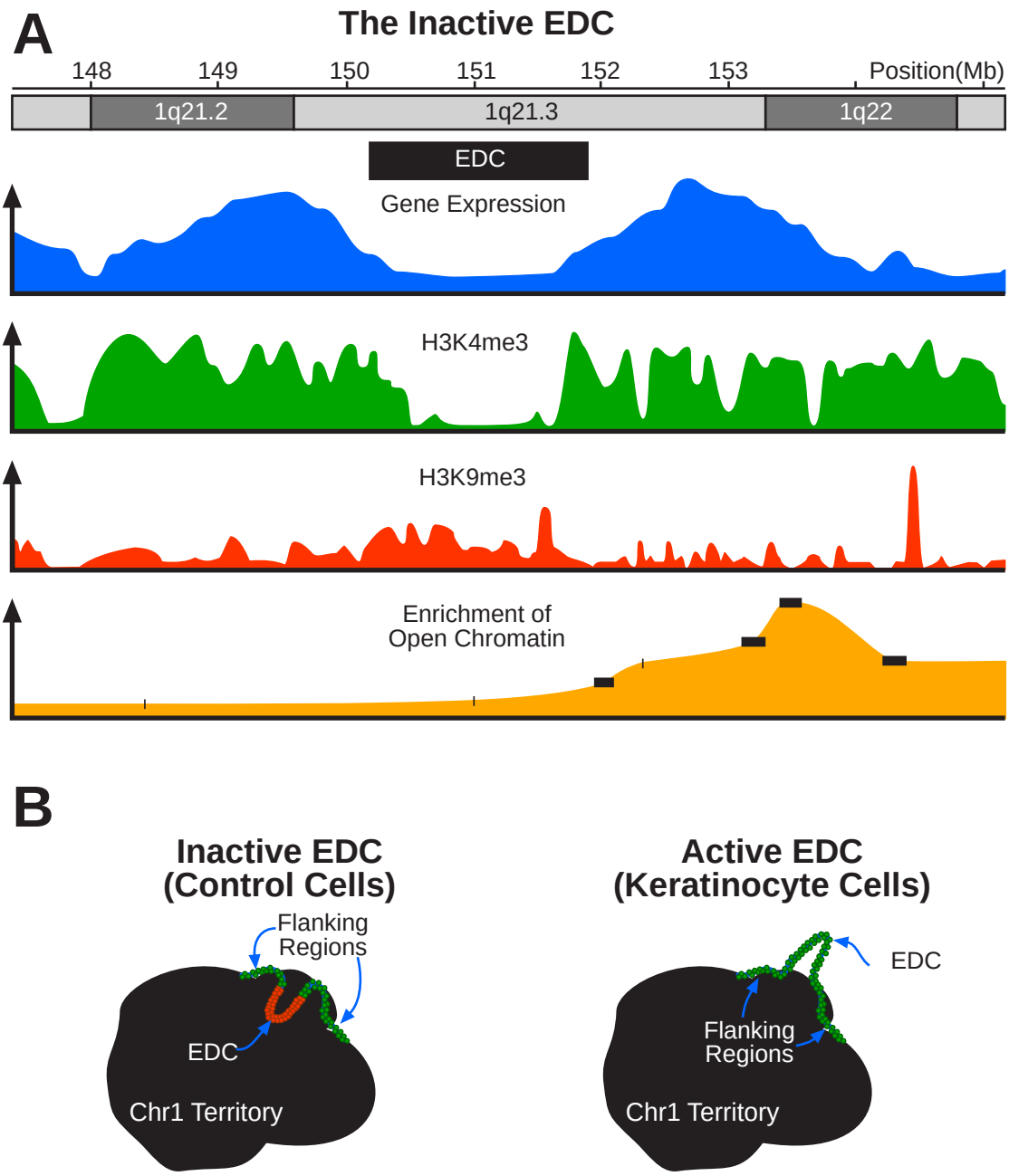


Figure 1.6 Hypothesis of Chromatin Structure at the EDC

A. Data from other studies of chromatin structure at the inactive EDC in Lymphoblastoid cells. In all cases data was roughly traced and so the graphs are not quantitative. Gene expression was measured from K562 lymphoblastoid cells and represents a moving median (Goetze *et al.*, 2007). H3K4me3 and H3K9me3 were traced from the UCSC annotation of the data of Barski *et al.*, 2007 and represents the results of mass sequencing of ChIP from $CD4^+$ human T-cells. The enrichment of open chromatin represents the hybridisation of biophysically isolated open chromatin fibres from FATO lymphoblastoid cells (Gilbert *et al.*, 2004). This data was sampled at a much lower resolution than the other datasets, on a BAC clone array. The positions of the BACs are indicated and they are drawn to scale. All genomic positions are from NCBI Build 36.

B. Hypothesis for the active EDC. Combining these data and the FISH data from other studies (Goetze et al., 2007, Williams et al., 2002) I propose that in active EDC will relocate outside of its chromosome territory and chromatin fibre structures across the region will adopt a more active conformation, the repressive histone marks will also be replaced with activating marks. Only one chromosome territory is shown in each case

Chapter 2: Materials and Methods

Unless otherwise stated, all reagents used were of analytical grade and obtained from Sigma or Invitrogen

2.1 Reagents, Stock Solutions and Buffers

Note: A number of solutions were available as stock solutions, prepared by technical services at the MRC Human Genetics Unit. These are indicated below and I am indebted to them for this service. All solutions were made using deionised water unless otherwise stated.

Alkaline Lysis Solutions: Solution I was 50mM Glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0). Solution II was 0.2N NaOH and 1% (w/v) SDS and was freshly prepared before use. Solution III was prepared by the mixing of 60ml of 5M Pottasium Acetate, 11.5ml of Glacial Acetic acid and 28.5ml of water (Sambrook and Russell, 2001).

Antibiotics: Stock solutions of chloramphenicol were made by dissolving chloramphenicol powder in ethanol to a concentration of 25mg/ml. Similarly kanamycin monosulphate was dissolved in water to a concentration of 10mg/ml to make a stock solution.

Buffered Phenol-Chloroform: Was a 25:24:1 mixture of Phenol, Chloroform and Iso-Amyl Alcohol saturated with 10mM Tris (pH 8.0) and 1mM EDTA and was purchased from Sigma.

ChIP Elution Buffer: 0.1M Na HCO_3, 1% SDS, was made fresh before use

Chromatin Wash Solutions: These solutions were used to prepare Chromatin for Crosslinking ChIP, they were all supplemented with 0.25mM PMSF

and Complete Protease Inhibitors (Roche). Chr-W1 was 10mM Hepes (pH 8), 10mM EDTA, 0.5mM EGTA, 0.25% Triton X-100 (v/v). Chr-W2 was 10mM Hepes (pH 8), 1mM EDTA, 0.5mM EGTA, 200mM NaCl, 0.01% Triton X-100 (v/v). Chr-W3 was 25mM Tris HCl (pH 8), 2mM EDTA, 150mM NaCl and 0.1% SDS (w/v).

Coomassie Stain: 0.25% Coomassie brilliant blue R-250 (w/v), 45% methanol (v/v) and 10% glacial acetic acid (v/v). The Coomassie powder was dissolved in the methanol first to avoid the formation of precipitates.

Coomassie Destain Solution: 30% methanol (v/v) and 10% glacial acetic acid (v/v).

Crosslinking Solution: 100mM NaCl, 50mM Hepes (pH 8) and 11% formaldehyde (w/v).

DNA Loading Buffer: 5xTBE with 40% Sucrose (w/v) and 0.25% Orange G (w/v)

Ficoll Loading Buffer: Was prepared by supplementing either 5xTBE or 5xTPE with 15% Ficoll (Type 400; Pharmacia, w/v) and 0.25% bromophenol blue (w/v) as appropriate.

FISH Hybridisation buffer: 50% deionised formamide (v/v), 10% dextran sulphate (v/v) 1% Tween 20 (v/v), in 2x SSC.

Genomic Lysis Buffer: Was prepared as a 2x stock solution: 300mM NaCl, 1% SDS (w/v) and 20mM EDTA. Samples were mixed at a 1:1 ratio (v/v) for use.

Luria-Bertani (LB) Broth: Was prepared by the addition of 10g of tryptone, 5g of yeast extract and 10g of NaCl to 1 litre of water. It was brought to pH 7.0 by the addition of Sodium Hydroxide. Was prepared by technical services at the MRC Human Genetics Unit.

Microarray Hybridisation Buffer: 50% Deionised formamide (v/v), 5xSSC, 5x Denhardt's solution and 0.1% SDS (w/v). The Denhardt's and SSC was purchased from Sigma to ensure it was free from nucleases and contaminants.

MOPs Electrophoresis Buffer: 20mM MOPS, 5mM Sodium Acetate, 1mM EDTA was made as a 10x stock using RNase free water.

Nuclear Buffer A (NB-A): 85mM KCl, 10mM Tris-HCl (pH 7.6), 5.5% Sucrose (w/v), 0.5mM Spermidine, 0.2mM EDTA, 0.25mM PMSF. Was made without PMSF and stored at 4^{o}C until use, upon which PMSF was added.

Nuclear Buffer R (NB-R): 85mM KCl, 10mM Tris-HCl (pH 7.6), 5.5% Sucrose (w/v), 0.25mM PMSF. Was made without PMSF and stored at 4^{o}C until use, upon which PMSF was added.

Nuclear Buffer ChIP (NB-ChIP): 85mM NaCl, 10mM Tris-HCl (pH 7.6), 5.5% Sucrose (w/v), 0.25mM PMSF, 5mM Na Butryate. Was made without PMSF and Na Butyrate and stored at 4^{o}C until use, upon which PMSF was added.

Nuclear Buffer ChIP (NB-Stop): 215mM NaCl, 40mM Tris-HCl (pH 8), 20mM EDTA, 5.5% Sucrose (w/v), 0.2% Triton X-100 (v/v), 0.25mM PMSF, 5mM Na Butryate, 2x Complete Proteinase Inhibitors (Roche). Was made fresh before use.

Phosphate Buffered Saline (PBS): Dulbecco's PBS (without Ca^{2+} and Mg^{2+}) was 10mM Phosphate, 137mM NaCl and 27mM Potassium Chloride. Was made from tablets purchased from Unipath (Oxford) by technical services at the MRC Human Genetics Unit.

Proteinase K: Lyophilized powder was dissolved to a concentration of 20mg/ml (w/v) in 50mM Tris (pH 8.0), 1.5mM Calcium Acetate, 50% Glycerol (v/v) and was stored at -20^{o}C.

RNA Loading Buffer: Was made as a 5x stock: 5xMOPS electrophoresis buffer with 4mM EDTA, 0.89M formaldehyde, 20% (v/v) glycerol, 30% Formamide (v/v) and 0.25% Bromophenol Blue (w/v) using RNase free water.

SDS: Was prepared as a 20% (w/v) stock solution by technical services at the MRC Human Genetics Unit.

SDS Loading Buffer: Was prepared as a 2x stock solution of 125mM Tris HCl (pH 6.75), 4% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 20% (v/v) glycerol and 0.1% (w/v) bromophenol blue.

SDS-Page Running Buffer: Tris-glycine running buffer: 25 mM Tris base, 250 mM glycine (pH 8.3) and 0.1% SDS (w/v) was prepared as a 5x stock (Sambrook and Russell, 2001).

SSC: 3M NaCl, 0.3M tri-sodium citrate, pH7.4 was prepared as a 20x stock by technical services at the MRC Human Genetics Unit.

STE: 10mM Tris-HCl (pH 8.0), 0.1M NaCl and 1mM EDTA (pH 8.0)

TE: 10mM Tris HCl (pH7.6), 0.1mM EDTA prepared by technical services at the MRC Human Genetics Unit.

TEEP Buffers: Were 10mM Tris HCl (pH 7.6), 0.1mM EDTA, 250μM PMSF and a salt concentration as indicated by the suffix, eg. $TEEP_{20}$ was made with 20mM NaCl. Prepared as a 20x stock and PMSF added before use.

Tris Borate Buffer, TBE: 90mM Tris Borate, 2mM EDTA (pH 8.0) was prepared as a 20x Stock Solution by dissolving 108g of Tris Base, 27.5g of Boric Acid in 40ml of 0.5M EDTA and 960ml of water and was diluted before use. Was prepared by technical services at the MRC Human Genetics Unit.

Tris Buffered Saline (TBS): 150mM NaCl and 10mM Tris-HCl (pH 7.4) was normally prepared as a 10x stock.

Tris Phosphate Buffer, TPE: 90mM Tris-Phosphate, 2mM EDTA (pH 8.0) was preoared as a 10x Stock Solution by dissolving 108g of Tris Base in 15.5ml of 85% phosphoric acid, 40ml of 0.5M EDTA and 944.5ml of water and was diluted before use.

Western Transfer Buffer: 25mM Tris-Glycine (pH 8.3) with 20% methanol (v/v) was made fresh before use. 3.03g of Tris Base and 14.4g of Glycine were dissolved in 800ml of water and 200ml of Methanol was added.

X-ChIP sonication buffer: 25mM Tris HCl (pH 8.0), 2mM EDTA, 150mM NaCl, 0.5% SDS (w/v) supplemented with 0.25mM PMSF and Complete Proteinase Inhibitors (Roche).

2.2 Plasmids and Bacterial Culture

2.2.1 Genomic Clones

All genomic clones were supplied by the Sanger Institute Clone service. Fosmids were from the WIBR2 library derived from a human female at the Broad Institute and had been cloned into DH10B T1 resistant competent cells. The vector backbone used for the derivation of the library was pEpiFOS-5. The BACs used were from the RPCI-11 Human Male BAC Library (Osoegawa et al., 2001). The cloning vector used was the pBACe3.6 vector and they were transformed into DH10B electrocompetent cells. The single PAC used in this study comes from the RPCI-1 Human male PAC library which was constructed by P.A. Ioannou at the Children's Hospital Oakland Research Institute using the pCYPAC2 vector and transformed into DH10B electrocompetent cells.

The stocks of these clones were all supplied as bacterial stab cultures in agar. These were then streaked out onto agar plates with an appropriate selective antibiotic and grown overnight at 37^{o}C to give single colonies. To ensure each that each stab was clonal, DNA was prepared from two of these colonies (see below) and digested using the enzyme EcoR1. By resolving the restriction fragments from each clone on agarose gels and comparing their patterns, all supplied clones were found to be clonal.

2.2.2 Bacterial Culture

Bacteria were streaked out onto plates made from Luria-Bertani (LB) broth supplemented with 1.5% agar (Prepared by technical services at the MRC Human Genetics Unit). The LB-agar was melted and selective antibiotics added after it had cooled, before the plates were poured. Bacteria were then streaked out to give single colonies and left to grow overnight at 37^{o}C. For BAC and Fosmid clones, plates

were supplemented with 37.5 μg/ml of chloramphenicol and PACs were grown on plates supplemented with 10 μg/ml kanamycin.

To prepare DNA from Bacteria they were grown in overnight liquid cultures. A single colony was picked from an agar plate and inoculated into 5ml of LB broth. They were then allowed to grow overnight with shaking at 37^{o}C and at least a 5:1 air to liquid ratio. The culture was also supplemented with the appropriate selective antibiotic, 12.5 μg/ml of chloramphenicol or 10 μg/ml kanamycin.

2.2.3 Bacterial Glycerol Stocks

To prepare a glycerol stock, glycerol was added to a final concentration of 30% (v/v) to an aliquot of an aliquot of an overnight culture. It was then stored at -70^{o}C.

2.2.4 Preparation of DNA from Bacterial Overnight Cultures

To prepare DNA from overnight cultures of genomic clones an adaptation of the method described by Sambrook and Russell was used (Sambrook and Russell, 2001). Firstly the overnight culture was spun down at 2000g for 5 minutes at 4^{o}C. It was then resuspended in ice cold STE buffer before being spun again and being resuspended in 200μl of cold alkaline lysis solution I on ice. 400μl of fresh alkaline lysis solution II was then added; the tube was inverted gently several times and again placed on ice. 300μl of ice cold alkaline lysis solution was then added, the tube inverted again and incubated on ice for 5 minutes. Precipitated cell debris were then pelleted by spinning at maximum speed in a microcentrifuge for 5 minutes at 4^{o}C and the supernatant removed to a fresh tube. Then two Buffered Phenol-Chloroform (BPC) extractions and one chloroform extraction were performed (see *Section 2.3.1* below) before RNase A was added to 100μg/ml and the preparation was incubated at 37^{o}C for 1 hour. A further BPC extraction and Chloroform extraction were then performed and 900μl of isopropanol was added and the DNA precipitated by

spinning for 15 minutes at maximum speed in a microcentrifuge. The pellet was then washed twice with 70% (v/v) ethanol before being left to resuspend overnight at 4^{o}C in 15-30µl of water. Generally these preparations yielded ~1µg of DNA (as determined on an agarose gel, spectrophotometry yielded inaccurate results).

2.3 Preparation and Handling of DNA

2.3.1 Purification of DNA by Phenol/Chloroform Extraction and Ethanol Precipitation

If necessary the volume of the sample was adjusted to at least 200µl with water. If substantial amounts of protein were present (for example in a nuclear or chromatin preparation) SDS was added to 0.1% (w/v), proteinase K to 20µg/ml and the preparation incubated at 55^{o}C for at least 1hr. An equal volume of BPC was added to the DNA sample, mixed, and centrifuged at 12,000 g for 5 minutes. After centrifugation the top aqueous layer was removed, avoiding the white precipitate present at the boundary between the two layers, and added to a fresh tube. For crude preparations of DNA this process was then repeated. Then an equal volume of pure chloroform was then added, and the mixing and centrifugation repeated to remove any remaining phenol.

To concentrate the DNA ethanol precipitation was performed as follows. 1/10th volume 5M Sodium Acetate at pH4.6 and 2 volumes of ice cold ethanol were added to the DNA sample. The sample was then mixed and incubated on ice for at least 15 minutes before being centrifuged at 12,000 g for 15-30 minutes, the supernatant discarded and the pellet washed in 70% ethanol (v/v). Centrifugation was repeated for 5 minutes and the supernatant removed. The pellet was then dried briefly at room temperature to remove excess ethanol and resuspended in the appropriate volume of water or other buffer as stated in the text.

2.2.3 Isolation of Genomic DNA

Genomic DNA was isolated by diluting nuclei prepared from cell lines into Genomic Lysis buffer. They were then briefly sonicated to shear the DNA before being cleaned up by Proteinase K digestion, two BPC extractions, a Chloroform extraction and precipitated with ethanol before being allowed to resuspend overnight at 4^{o}C.

2.2.3 Digestion of DNA with Restriction Endonucleases

For analysis of genomic clones, DNA was digested using the EcoRI enzyme. 100-200ng of DNA was diluted in the appropriate buffer supplemented with 100μg/ml Bovine Serum Albumin (BSA) and digested with 20 Units of EcoR1 (New England Biolabs) by incubation for 1hr at 37^{o}C. The reaction was stopped by the addition of DNA Loading Buffer (see below) and the products resolved on a 0.7% agarose gel.

2.3.4 Quantitation of DNA by Spectrophotometry

DNA concentration was routinely measured spectrophotometrically using a nanodrop ND-1000 (Nanodrop Technologies) according to the manufacturer's instructions. The absorbance of the sample was measured at a wavelength of 260nm (A_{260}). An A_{260} of 1 equals 50μg/ml of double stranded DNA.

2.3.5 Resolution of DNA on Agarose Gels

Gels for routine analysis of DNA were made to an appropriate agarose percentage ("Hi-Pure" Low Eeo agarose, BioGene UK, w/v) in Tris-Borate Buffer (TBE). DNA

Loading buffer was added to all samples at a ratio of 1 part buffer to 4 parts sample. Appropriate commercial DNA size markers were used to allow size determination and quantification of DNA fragments. RT PCR products were resolved on 4% gels made with NuSieve 3:1 agarose (Cambrex Bioscience Rockland, Inc). Preparations of Mono-Nucleosomes were resolved on 4% gels made with NuSieve GTG agarose (FMC Bioproducts, Rockland USA). DNA prepared from chromatin fractionated on sucrose gradients was resolved on 0.7% agarose gels using Tris-Phosphate (TPE) Buffer. Ficoll DNA loading buffer prepared with TPE buffer was also used. These were run overnight and the buffer recirculated. TPE was used as a buffer because it is thought to give better resolution of high molecular weight DNA fragments (N. Gilbert, PhD theis, University of Edinburgh). After resolution the gels were stained in electrophoresis buffer with 0.5 mg/ml Ethidium Bromide for at least 30 minutes at room temperature and destained in electrophoresis buffer for 15 minutes at room temperature. DNA was then visualised under Ultra Violet illumination. For analysis gels were scanned using a 532nm laser and a 575nm Long Pass filter (LPG filter on a Fuji FLA-5100 phosphorimager).

2.3.6 Pulsed Field Gel Resolution of DNA on Agarose Gels

Preparative fractionation of DNA from gradient fractions was carried out by pulsed-field gel electrophoresis (PFGE) (CHEF system, Biorad) through 1% Seaplaque GTG, low melting point agarose (FMC Bioproducts) in 0.5 × TBE, at 180 V, for 36 hr, with a 0.1–2 s switching time. Size markers were 1 kb (Promega) and 2.5 kb (Biorad) DNA ladders.

2.3.7 Extraction of DNA Fragments From Agarose Gels

DNA was extracted from gels using the β-Agarase I Enzyme (New England Biolabs). Transverse gel slices containing the DNA of interest were cut using a clean razor blade and melted in a microcentrifuge tube at 65^{o}C. The volume of melted gel was then estimated before it was diluted into 1x β-Agarase I Reaction Buffer (10mM Bis-Tris HCl, 1mM EDTA pH 6.5) and incubated at 65^{o}C for 10mins. It was then cooled to 42^{o}C before an appropriate amount of enzyme was added (1 unit of β-Agarase will digest 200µl of 1% low melting point agarose in 1 hour) and incubated at 42^{o}C for 2 hours. After this period, the sample was cooled on ice and examined to make sure all agarose had been digested. Two BPC extractions and one chloroform extraction were then performed before the samples were ethanol precipitated and resuspended in an appropriate volume of water.

2.3.8 Amplification of DNA Samples by Whole Genome Amplification

The yield of 'open' chromatin DNA that could be obtained from preparative pulse field gels was very low. Therefore, before hybridisation to microarrays it had to be amplified. I chose to use a commercial kit, the GenomePlex Whole Genome Amplification Kit (WGA) kit from Sigma for this purpose. The kit representatively amplifies DNA from as little as 10ng. Briefly the DNA is fragmented chemically before linkers are ligated to the fragments and the DNA is amplified by PCR.

Samples were first diluted to a concentration of 1ng/µl in water, then 1µl of 10x Fragmentation Buffer was added to 10µl of the dilute sample in a microcentrifuge tube and mixed. The tube was placed on a thermal block at 95^{o}C for exactly 4mins to fragment the DNA before being cooled on ice. Then 2µl of Library Preparation Buffer and 1µl of Library Stabilization solution were added, mixed and the tube heated at 95^{o}C for 2mins and cooled on ice. 1µl of Library

preparation enzyme was then added and the following series of incubations performed on a thermocycler: 16°C for 20mins, 24°C for 20mins, 37°C for 20mins, 75°C for 5mins. Then the samples were cooled on ice before amplification. 7.5µl of 10x Amplification Master Mix, 47.5µl of water and 5µl of WGA DNA Polymerase were added, mixed and the following program of thermocycling performed: 95°C for 3 minutes, followed by 14 cycles of 94°C for 15secs and 65°C for 5mins. After amplification, the samples were cleaned up by performing one BPC extraction and one Chloroform extraction. After ethanol precipitation they were resuspended in 50µl of 10mM Tris-HCl (pH 8.0).

For the amplification of DNA prepared for 'open' chromatin, it was assumed that the concentration of DNA recovered from gel slices was roughly 1ng/µl. When the amplified DNA was resolved on a an agarose gel, it was found that, as expected it ranged in size from around 75-1500bp and no DNA was present in a negative control sample prepared using water instead of DNA (*Figure 2.1A* and data not shown). It was found that two preparations of 'open' chromatin were more dilute than expected. These samples were amplified for 20 cycles rather than 14 (the corresponding input samples were amplified similarly). No product was observed in a similarly amplified negative control and after two technical replicate array hybridisations these samples were correlated to two independent samples similarly treated but amplified for only 14 cycles (Pearsons, R^2: 0.523 for the HaCaT sample and 0.549 for the 293 sample).

I also checked whether this methodology was indeed capable of representatively amplifying a complex mixture of fragments. My experiments using Chromatin Immunoprecipitation (ChIP) had revealed that sufficient DNA could be precipitated to allow hybridisation to oligonucleotide microarrays without prior amplification (see *Chapter 5*). I therefore performed ChIP from HaCaT cells using an anti-Acetylated Histone H3 Lysine 9 (H3K9Ac) antibody (see *Section 2.8*, below). Two technical replicate hybridisations of labelled ChIPed material and Input material were then performed on my custom oligonucleotide microarrays (see *Section 2.9*, below). Aliquots of the remaining ChIPed and Input material were then amplified using the WGA method described above. The amplified material was then used to perform two technical replicate hybridisations. Both sets of hybridisations were scanned, quantified and normalised using Bluefuse. A mean $\log_2$

H3K9Ac/Input ratio was generated for each oligo for both the amplified and unamplified material. The correlation between amplified and unamplified mean $\log_2$ values was 0.459 which is lower than that observed between two independent biological replicates of the ChIP (R^2, Pearson's = 0.762, *Figure 2.1B*). However, I was able to detect the enrichment of H3K9Ac at the active *S100A* gene promoters in the amplified material like the unamplified material (*Figure 2.1C*). Therefore, the kit representatively amplifies complex mixtures of DNA.

2.3.9 Labelling of DNA Samples for Hybridisation to Microarrays

DNA samples were labelled for hybridisation to microarrays using Bioprime Total kits (Invitrogen) according to the manufacturer's instructions. These kits use nucleotides conjugated to Alexa Fluor 3 and 5 dyes rather than the more commonly used Cyanine 3 and 5. However, for the purposes of detection, they are essentially identical to the Cyanine dyes. Therefore, throughout this text I will refer to Alexa Fluor 3 and 5 as Cyanine 3 and 5. Briefly 50-500ng of DNA were diluted into 1mM EDTA and the primer/ labelled nucleotide mix added. The DNA was then denatured at 95^{o}C for 5mins and cooled on ice for 5mins before the Exo^{-} Klenow enzyme fragment was added. The labelling reaction was allowed to proceed for 3hrs at 37^{o}C with mixing and was stopped by the addition of 5µl 0.5M EDTA. This was one hour longer than given in the manufacturer's instructions but the extra hour was found to increase the labelling efficiency (data not shown).

The labelled DNA was then purified using QIAGEN PCR purification columns. These were used according to the manufacturer's instructions but with some modifications in order to increase the efficiency of the procedure. The sample mixed with buffer PBI was passed twice over the column in order to increase the efficiency of binding. Elution was performed by adding 25µl of water (made alkaline by the addition of a drop of NaOH) to the column and then incubating the column for 5mins at RT. The elution was then repeated to ensure the whole sample was eluted. Samples and reagents were kept in the dark at all stages.

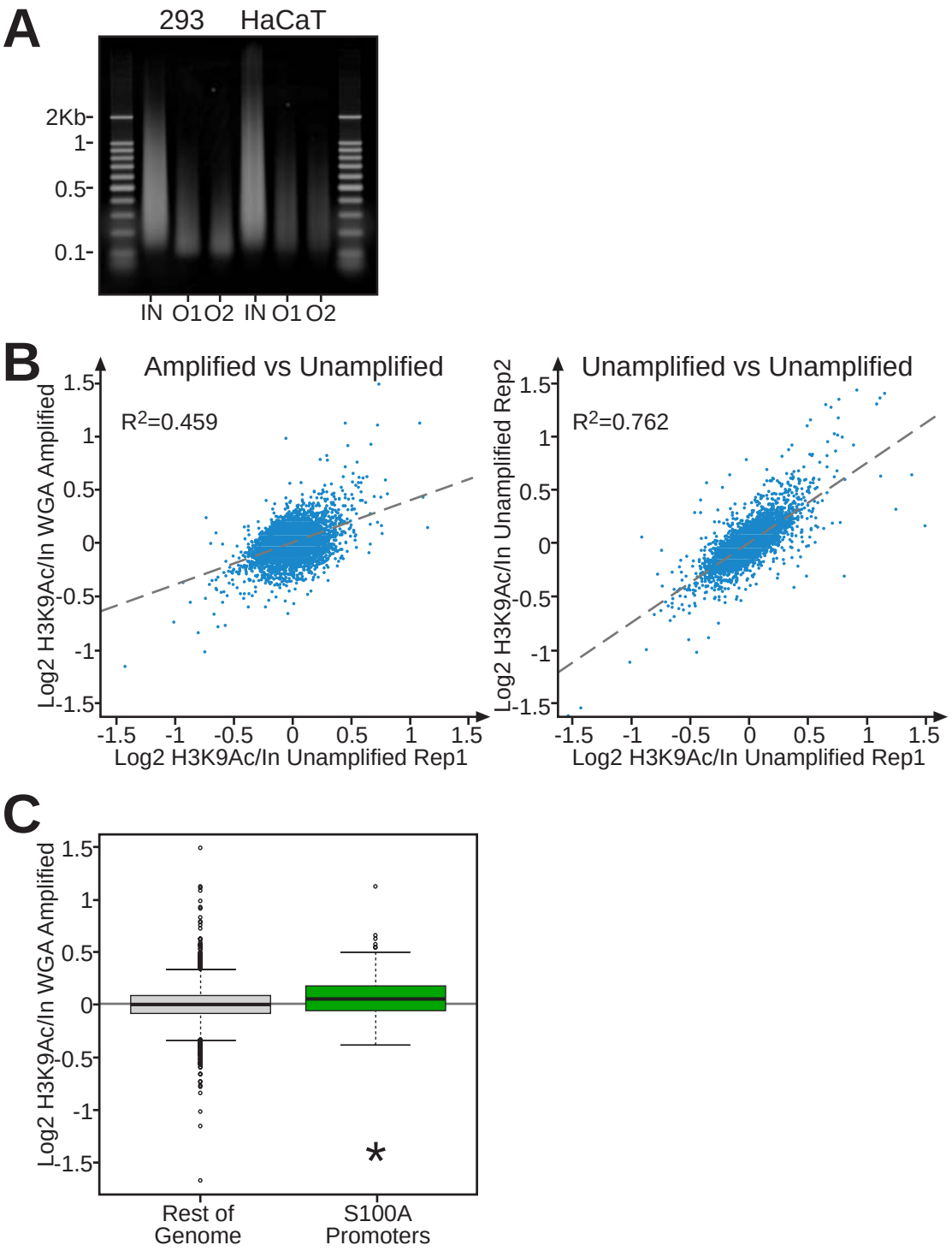


Figure 2.1 Testing the WGA Amplification Method

A. Amplified chromatin fractions from 293 and HaCaT cells. IN- Input, O1 and 2- Different Open chromatin fractions (see *Chapter 5*). Selected DNA size markers are indicated.

B. Scatter plots of Log2 H3K9Ac ChIP/Input for unamplified material vs amplified material and unamplified material from two biological replicates of the ChIP. The R^2, Pearson's is given and a line of best fit drawn.

C. Boxplot showing that enrichment of H3K9Ac at the promoters of the active S100A genes can be detected using amplified samples. The two distributions are significantly different (KS-test). A similar enrichment can be detected in unamplified material (see *Chapter 5*).

2.3.10 Real Time PCR

Real Time PCR analysis of DNA prepared from ChIP samples was performed with a Quantitect SYBR Green detection kit (Qiagen) according to the manufacturer's instructions, with a Peltier PTC-200 thermocycler using an in built Cromo4TM continuous fluorescence detector connected to Opticom 3.1 software interface. Standards were created from a serial dilution of ChIP input DNA using a 10% input aliquot (0.2%, 0.04%, $8x10^{-3}$% and $1.6x10^{-3}$% of input). 2µl of these standards and 2µl (4%) of precipitated and mock precipitated samples were then diluted into the reaction mixture along with 0.5nmol/ml of each primer. The following thermal cycling program was then followed: 15min at 95^{o}C followed by 45 cycles of 94^{o}C for 30secs, 66^{o}C for 30secs, 72^{o}C for 30secs. The fluorescence of each sample was read after each cycle at a temperature of 72^{o}C. Fluorescence readings were used to calculate the cycle at which each sample and standard crossed a fluorescence threshold (C_t). The C_t values of the standards were used to plot a graph of $\log_{10}$ input percentage against threshold cycle. The equation of this graph, calculated by linear regression, then allowed a percentage of input DNA value to be assigned to the precipitated and mock precipitated samples. The precipitated samples were then expressed by their degree of enrichment over the mock sample (ie %input of ChIP/%input of Mock).

The primers used were specific for the *S100A6* gene promoter region and were GCGCAGCTCAGTTTGCTCAGTTT (forward) and TCACATTCACGGTCTGGCAGAGGAA (reverse), see *Figure 5.16*, p186. Using a melting curve from 60 to 100^{o}C and by resolution on an agarose gel, I confirmed that the amplification conditions used with these primers produced a single product of the correct size 111bp).

2.4 Preparation and Handling of RNA

2.4.1 RNA isolation and purification

To isolate and purify total RNA from cells, I used TRI-reagent (Sigma) according to the manufacturer's instructions. Briefly, adherent cells were washed in PBS at room temperature before ~100µl of TRI-reagent were applied per cm^2 of tissue culture flask. Cells grown in suspension were pelleted at 1000g for 4mins at RT before being washed once in PBS and resuspended in a small volume of PBS and an appropriate volume of TRI-reagent added as above. The cells were incubated a few minutes in the TRI-reagent to lyse them before being transferred to an appropriate vessel. Cells were then homogenised and genomic DNA sheared by passing the suspension through a 21 gauge needle. Samples were then left at 4^oC for 5mins to allow complete dissociation of nucleoprotein complexes. Samples could be frozen at -70^oC at this stage for later use and then thawed before being processed. $1/5^{th}$ of a volume of chloroform was then mixed with the TRI Reagent, before being centrifuged at 12,000g at 4^oC to separate the two phases. The RNA is found in the upper aqueous phase of the homogenate. This is removed and 0.5 volumes of isopropanol were added before being incubated for 5mins at RT and pelleted by centrifugation at 12000g for 15 minutes at 4^oC. The pellet was washed twice in 75% ethanol (v/v) and could be stored at -70^oC in 75% ethanol. After the final wash, excess ethanol was removed by brief air drying and the pellet resuspended in an appropriate volume of RNase free water. RNA was quantified by measuring the A_{260} and concentration calculated as described in Sambrook et al. (2001).

2.4.2 Purification of RNA by Phenol/Chloroform Extraction and Ethanol Precipitation

Buffered Phenol-Choloroform was used to clean up RNA samples in the same way as for DNA samples (see *Section 2.3.1*, p56). Only one BPC extraction was

performed on RNA samples. To concentrate the RNA, ethanol precipitation was performed as for DNA, but the pellet was washed using 75% ethanol (v/v) rather than 70%. The pellet was then dried briefly at room temperature to remove excess ethanol and resuspended in the appropriate volume of RNase free water or other buffer as stated in the text.

2.4.3 Resolution of RNA on Formaldehyde Gels

1.2% agarose gels ("Hi-Pure" Low Eeo agarose, BioGene UK) were made in 1x MOPS electrophoresis buffer supplemented with 0.13 M formaldehyde. The gels were run in MOPS electrophoresis buffer supplemented with 0.25 M formaldehyde. The RNA samples were diluted into 1x RNA Loading buffer and denatured at 65^{o}C for 5mins and cooled on ice for 5mins before loading. During electrophoresis the buffer was recirculated using a peristaltic pump. After resolution the gels were stained in 0.5 mg/ml Ethidium Bromide for 30mins at RT and destained in 1x MOPS electrophoresis buffer for 15mins at RT. The RNA was visualised under UV illumination.

2.4.4 Reverse transcription polymerase chain reaction (RT-PCR)

20μg of RNA was treated with DNase I (0.3U/μl) and Protector RNase inhibitor (1.3U/μl, both Roche) for 1 hour at 37^{o}C and then phenol/chloroform extracted and ethanol precipitated as described above. Treated RNA was quantified by measuring its A_{260} and its integrity checked using a formaldehyde gel to compare it to the untreated RNA (see above). Reverse transcription to create first strand cDNA was then performed using random hexamers. 1-2μg of RNA was mixed with 200μM of random hexamers and incubated at 70°C for 10mins before being transferred to ice. To create the first strand cDNA, the RNA was incubated with 10mM dithiothreitol (DTT), 0.5mM dNTPs, 40U of Protector RNase inhibitor and 400U of Superscript II

reverse transcriptase (Invitrogen) at 37°C for 1hr in a total volume of 40µl. PCR was then performed with 2.5mM $MgCl_2$ using the first strand cDNA as a template and two primers (0.5µM each). The primers and PCR programmes used are described below. The DNA PCR products were resolved on 4% 3:1 NuSieve Agarose TBE gels as described in *Section 2.3.5*, p57.

2.4.5 Primers for RT-PCR

The primer pairs were designed to yield products of around 100 to 200 base pairs and to span an intron of their target mRNA, thereby allowing the presence of contaminating genomic DNA in PCR reactions to be identified. The *β-Actin* Primers were kindly donated by J. Sanford.

Table 2.1

Primer Pairs and conditions used for RT PCR

The following gene name abbreviations are used: *ILF2: Interleukin Enchancer-Binding Factor 2, SNX27: Sorting Nexin 27*, FLG: Filaggrin, *SPRR1B: Small Proline Rich Protein 1B, LCE3D Late Cornified Envelope Protein 3D, CK10* and *14: Cytokeratin K10* and *K14 respectively, IVL: Involucrin.*

Gene	Oligonucleotide Sequence	Product Size (bp)	PCR Conditions
β-Actin	GGGGGATCCCCGGCATGTGCAAGGCCGGC GGGGAATTCGCCGTGCTCGATGGGGTACT	178	10mins 95°C (30s 95°C, 30s 60°C, 1min 72°C)x25 3min 72°C
ILF2	TAAACGTTGCATACAGGCGCT GCT ACCATGTCCTGCTGTTCTAGGGTCAT	139	3mins 96°C (30s 96°C, 30s 60°C, 1min 72°C)x30 3min 72°C
SNX27	GTGTTCTGCGAGCTCAAGTGGAGA AGACAAGGGAAGAAGCAGGGCTTT	142	
S100A6	TGACAAGCACACCCTGAGCAAGAA TCCTTGTTCCGGTCCAAGTCTTCCAT	123	
S100A10	ATGCCATCTCAAATGGAACACGCC TTGTCCACAGCCAGAGGGTCTTT	161	
FLG	TGGGTCTGCTTCCAGAAACCATCA TGTTCATGGGATGATGCAGCCTGT	183	
SPRR1B	CACACCAGGACCAGCCACTGTTGCA TGACTATTGAAGGGCAGGGCTCAGG	259	
LCE3D	AGGAACAGCCTTCTCCTGCCTCCTC AGGCAGACACTGTACTGGGCTCTTTG	118	
CK10	TCCCAACTGGCCTTGAAACAATCCCT TGTTGCAACTGTTCTTCCAGAGCGGA	119	3mins 94°C (30s 94°C 30s 66°C 30s 72°C)x30
CK14	TCCTTCTGCAGATTGACAATGCCCGT ATTGATGTCGGCTTCCACACTCATGC	101	
IVL	GCCTCAGCCTTACTGTGAGT TGGGTATTGACTGGAGGAGGAACAGTC	137	

2.4.6 Preparation of Labelled cDNA for Hybridisation to Microarrays

40μg of RNA was treated with 4U DNase Turbo (Ambion) and Protector RNase inhibitor (1.6 U/μl) for 15 minutes at 37^{o}C. The RNA was then purified using RNeasy Min-elute columns (QIAGEN) according the manufacturers instructions. Briefly the RNA is selectively bound to a silica-gel membrane and washed before being eluted using RNase-free water. I chose to elute by adding 12μl of water to the column, incubating 5mins at RT and then spinning the column to elute the RNA. This process was repeated to ensure all the RNA was eluted. The RNA was then quantified and its integrity checked by resolution alongside the untreated sample on a formaldehyde agarose gel (see above). ~20μg of RNA was then labelled for hybridisation onto microarrays using the Superscript III direct labelling kit (Invitrogen) and Cyanine 3 and 5 labelled dCTPs (Amersham-Pharmacia). Briefly, the RNA was incubated with random primers for 5mins at 70°C and then cooled for 5mins on ice. RNA Guard, dNTPs, labelled dCTP and Superscript Enzyme were then added, as per instructions, and the mixture incubated at 46°C for 3hrs. To hydrolyse the RNA, 15μl of 0.1M NaOH were added and incubated for 15mins at 70°C, this was neutralised by the addition of 15μl 0.1M HCl and mixing. The labelled cDNA was then purified using a PCR purification kit (QIAGEN). Briefly, the labelled DNA is selectively bound to a silica-gel membrane and then washed to remove components of the reaction before being eluted using an alkaline buffer. I modified the manufacturer's protocol slightly by passing the DNA over the column twice to ensure maximal binding, washing twice and eluting after a 5min incubation at RT using 25μl of water (made alkaline by the addition of a drop of 5M NaOH to 50ml). The elution was also repeated twice to ensure maximal recovery of labelled DNA.

2.5 Preparation and Handling of Protein

2.5.1 Preparation of Whole Cell and Nuclear Protein Extracts

Whole cell protein extracts were prepared by washing cells in PBS and then adding an appropriate volume of 1xSDS loading buffer to lyse the cells. The solution was then transferred to a microcentrifuge tube and sonicated briefly to shear the chromatin. The samples were then boiled at 95^{o}C for 5-10mins to denature the proteins.

To prepare nuclear protein extracts, nuclei were prepared according to *Section 2.6.5*, below. They were then resuspended in a small volume of buffer and an equal volume of 2xSDS loading buffer was added and sonicated and boiled as for whole cell extracts.

2.5.2 TCA Precipitation of Proteins

To precipitate proteins, samples were mixed with an equal volume of 20% (v/v) Trichloroacetic Acid (TCA) and then stored for a minimum of 1hr at -20^{o}C. They were then spun for 15mins at maximum speed in a microcentrifuge at 4^{o}C. The supernatant was removed and the pellet washed twice with a large volume of ice cold, pure acetone. Between washes, the pellet was incubated for at least 30mins on ice and respun in a microcentrifuge. The second acetone wash was completely removed and the pellet dried on a heat block at 95^{o}C for 5-10mins to remove the last of the acetone. The pellet was then resuspended in an appropriate volume of 1xSDS loading buffer and boiled at 95^{o}C for 5-10mins.

2.5.3 SDS PAGE Resolution of Proteins

Cell protein extracts were resolved by Sodium Dodecylsuplphide Polyacrylamide Gel Electrophoresis (SDS-PAGE, (Sambrook and Russell, 2001)). Briefly, denaturing polyacrylamide minigels with 8-12% acrylamide (v/v), 0.39 M Tris-HCl (pH 8.8), 0.1% SDS (w/v), 0.1% ammonium persulfate (w/v) and 0.04% N, N, N', N-tetramethylethylene diamine (TEMED, v/v) were prepared in water and left to set covered with a butan-1-ol overlay. This was then removed and a stacking gel with 5% acrylamide (v/v), 0.13 M Tris-HCl (pH 6.8), 0.1% SDS (w/v), 0.1% ammonium persulfate (w/v), and 1% TEMED (v/v) applied on top. After this had set, the gel was placed in SDS-PAGE running buffer, the samples loaded and run in electrophoresis tanks (Mighty Small, Hoefer) at 110 V for ~2hrs. Pre-stained protein standards (Bio-Rad) were loaded alongside samples to aid with analysis. All acrylamide used was from a 30%, 29:1 acrylamide:bis-acrylamide (v/v) stock (Severn Biotech).

2.5.4 Visualisation of Resolved Proteins

Firstly stacking gels were removed from SDS-PAGE gels and the denaturating gel washed twice in water. Gels were then submerged in for 1 hour at RT with gentle agitation in Coomassie Stain Solution. Stain was then discarded and gels incubated overnight in Coomassie Destain Solution, again with gentle agitation.

2.5.5 Western blotting

After SDS-PAGE gels had been run (see above), the stacking gel was removed and protein samples transferred to polyvinylidene difluoride (PVDF) membrane (Hybond-P, Amersham-Pharmacia) using a Genie Blotter (Idea Scientific). In brief the membrane was soaked in methanol, followed by water and equilibrated in

Western Transfer Buffer. The apparatus was then assembled according to the manufacturer's instructions with the gel and membrane sandwiched between 3MM paper (Whatmann) and the proteins transferred 1hr.

To reduce background, membranes were blocked for 1hr at RT in TBS-Tw (1x TBS supplemented with 0.05% Tween 20) supplemented with 4% milk proteins (Marvel). Primary antibodies were then diluted to the appropriate concentration in TBS-Tw 4% milk and incubated with membranes at 4^{o}C overnight with constant agitation. The next day, membranes were washed 3 times for 5mins at RT in TBS-Tw before being incubated with secondary antibodies for at least 1 hour at RT in TBS-Tw 4% milk. Finally washes in TBS-Tw were repeated and membranes developed using enhanced chemiluminscent detection (SuperSignal Western Pico Reagent, Pierce). Signals were exposed on Hyperfilm Ecl (Amersham-Pharmacia). For primary antibodies used and their dilutions see *Table 2.2*, below. Secondary antibodies used were, ImmunoPure Goat Anti-Mouse IgG + IgM (H+L), Peroxidase Conjugated (Pierce) and Goat Anti-Rabbit IgG (Whole Molecule) HRP (Sigma) both at a dilution of 1:10,000.

Table 2.2

Primary Antibodies for Western Blotting

Histone modifications are abbreviated, H3K9Ac (Acetylated Histone H3 Lysine 9), H4K16Ac (Acteylated Histone H4 Lysine 16), H3K9Me2 (di-methylated Histone H3 Lysine 9).

Epitope	Supplier	Catalogue No	Species	Dilution
Involucrin (SY5)	Abcam	68-100	Mouse	1:5,000
Cytokeratin 10 (LH2)	Abcam	20208-100	Mouse	1:1,000
Cytokeratin 14 (RCK107)	Abcam	9220-100	Mouse	1:500
Histone H3	Upstate	07-690	Rabbit	1:50,000
H3K9Ac	Upstate	07-352	Rabbit	1:10,000
H4K16Ac	Upstate	07-329	Rabbit	1:5,000
H3K9me2	Upstate	07-441	Rabbit	1:1,000

2.6 Cell Culture

2.6.1 Cell lines

575 cells are male lymphoblastoid cells that are known to have a normal karyotype (Woodfine et al., 2004). 293 cells are a human embryonic kidney cell line that has been transformed by exposure to sheared fragments of human adenovirus type 5 DNA (Graham et al., 1977). However, a microarray expression analysis has suggested that they are neuronal in origin (Shaw et al., 2002). HT1080 are a human fibrosarcoma cell line derived by Rasheed and collegues (Rasheed et al., 1974). HaCaT cells are a spontaneously immortalized keratinocyte cell line (Boukamp et al., 1988). SVK14 and, NEB1 cells are human keratinocytes that have been transformed by SV40 infection and, the early region of human papillomavirus 16 respectively (Morley et al., 2003;Taylor-Papadimitriou et al., 1982). HaCaT, SVK14 and, NEB1 were kindly supplied by of B. Lane. A second clone of the HaCaT cell line was kindly supplied by N.E. Fusenig.

2.6.2 Thawing Cells from Storage

Aliquots of cell suspensions were stored in liquid nitrogen in cell culture media supplemented with 10% dimethyl sulfoxide (DMSO). After thawing at 37^{o}C, cells were seeded into an appropriate tissue culture flask. In order to remove the DMSO, the media was aspirated from the cells and replaced with fresh media after the cells had adhered to the flask.

2.6.3 Culture of Cell Lines

All cells were cultured at 37^{o}C with 5% CO_2 and tissue culture medium containing penicillin (10,000 units/ml) and streptomycin (650μg/ml). HT1080, HEK293, HaCaT and, SVK14 cells were cultured in high glucose Dulbecco's Modified Eagle

Medium (DMEM), supplemented with 10% (v/v) Foetal Calf Serum (FCS). NEB1 cells were cultured as described (Morley et al., 2003), in DMEM with 25%(v/v) Ham's F12 medium, 10% FCS, 0.4μg/mL hydrocortisone (Sigma), 5μg/mL transferring (Sigma), $2x10^{-11}$mol/L lyothyronine (Sigma), $1.9x10^{-4}$mol/L adenine (Sigma), 5μg/mL insulin and, 10ng/mL mouse recombinant EGF (BD Biosciences). 575 cells were cultured in Roswell Park Memorial Institute-1640 medium supplemented with 10% FCS, Sodium Pyruvate (1mM), non-essential amino acids (Sigma), Insulin, 2mM L-Glutamine, and MOPS (a few drops of 1.25M).

Cells were grown to near confluency before being passaged into new culture flasks by the following methods. To split HT1080, 293 and, SVK14 cells, spent tissue culture medium was removed by aspiration. Cells were then washed once in PBS before a small volume of trypsin-EDTA (10%, v/v) was added and incubated with the cells for 5 minutes at 37^{o}C. Cells were dislodged by simple agitation, medium added to stop the trypsin, and cells pelleted at 1000g for 4 minutes before replating. HaCaT and NEB1 cells were split in a similar manner except they were washed twice with PBS and then incubated for 5 minutes at 37^{o}C in small volume of trypsin-EDTA (15%, v/v). This trypsin-EDTA wash was then removed and a further small volume added for a second 5 minutes at 37^{o}C. The cells were then dislodged and replated as above. After replating NEB1 cells were grown for 24hrs in medium lacking EGF before this was changed for their normal EGF supplemented medium (Morley et al., 2003). 575 cells grow in suspension and therefore were pelleted as above and resuspended in fresh culture medium before being seeded into fresh flasks at an appropriate density.

2.6.4 *In-Vitro* Differentiation of HaCaT Cells

HaCaT cells were differentiated in vitro using a published protocol (Paramio et al., 1998). Briefly, the cells were plated at ~60-80% confluency and left to grow for 16-24 hours. Their medium was then replaced with DMEM containing antibiotics as normal but lacking FCS. Serum free medium was replaced with fresh medium every two days during the differentiation process. For time course experiments, batches of cells were aliquoted onto plates in parallel and the differentiation started on the same

day. The time point at which a sample plate was removed was defined as the number of days that had elapsed since the normal cell culture medium was replaced by serum free medium.

2.6.5 Preparation of Nuclei from Cell Lines

Cell nuclei were prepared using a previously described modification the method of Cereghini and Yaniv (Cereghini and Yaniv, 1984;Gilbert et al., 2003). Cells were harvested using trypsin-EDTA (10%, v/v) and washed with PBS at RT. The cell pellet was then resuspended in a small volume of Nuclear Buffer A (NBA) before being diluted to 5ml volume. An equal volume of NBA supplemented with 0.1% Igepal-360 (v/v) was then added and the cells incubated on ice for 3mins, during which time the cell membrane lyses leaving the nucleus intact. The nuclei were collected by centrifugation (1,200rpm for 4 minutes at 4^{o}C in a benchtop centrifuge). The successful production of nuclei could be monitored by a change in colour of the pellet from off-white to white at this stage. Nuclei were then washed in Nuclear Buffer R (NB-R, essentially NB-A without EDTA). They were then resuspended in a small volume of NB-R (except nuclei for ChIP which were resuspended in NB-ChIP, see *Section 2.8.7*, below).

Nuclei concentration was determined by diluting a small aliquot at 1:20 in NB-R and adding a small amount of DNase I (Roche). The sample was incubated at 37^{o}C for ~5-10mins and further diluted into sonication buffer. The absorbance at a wavelength of 260nm was then measured using a Ultrospec 3000pro, Amersham Pharmacia Biotech. From this, DNA concentration was calculated as for *Section 2.3.4* (p57). Nuclei concentrations were expressed by the concentration of DNA they contained.

2.7 Fluorescence *In-Situ* Hybridisation (FISH)

2.7.1 Preparation of human chromosomes for 2D FISH

To increase the number of mitotic cells, and therefore the amount of chromosome spreads present, KaryoMAX colcemid (0.1μg/ml) was added to cell cultures for 30mins before harvest. Colcemid inhibits mitosis by depolymerizing microtubules. After 30 minutes of incubation cells were harvested, washed once in PBS (phosphate-buffered saline), and fixed in 3:1 methanol:acetic acid (MAA) for 2D FISH as described below.

2.7.2 Harvesting and Fixing Cells in 3:1 Methanol:Acetic Acid

Cells for two-dimensional (2-D) FISH were harvested, washed once in PBS at RT and resuspended in 10ml hypotonic solution (33mM KCl and 17mM Tri-Sodium Citrate). This was added drop-wise with constant agitation (the concentration of cells in hypotonic should be $<2x10^7$/ml). The cells were then left to swell for 10 min at RT before centrifugation at 400g for 5 min. The hypotonic solution was then aspirated off the nuclei pellet and cells were fixed in fresh 3:1 MAA. To do this, the first 2ml of fix was added dropwise with constant agitation, followed by the addition of a further 8ml of fix. This was repeated another two times. Cell preparations were then stored at -20^oC indefinitely.

2.7.3 Preparation of FISH probes

The analogues biotin-16-dUTP and digoxigenin-11-dUTP (Roche) were used to label all DNA probes. These analogues were incorporated into DNA by nick translation.

Following incorporation, unbound nucleotides were removed and the efficiency of labelling measured.

To label genomic clones for *in situ* hybridisation, 1-1.5μg of DNA was added to 4μl of nick translation salts (0.5M Tris-HCl pH7.5, 0.1M $MgSO_4$, 1mM DTT, 500μg/ml BSA), alongside 5μl each of 0.5mM dATP, dCTP and dGTP. Then either 5μl of 1mM biotin-16-dUTP or 3μl of 0.5mM dTTP along with 2μl of 1mM digoxigenin-11-dUTP was added to label the DNA. Freshly diluted DNase I (Roche) was then added to the translation to create a final concentration of 1 U/ml, alongside 1μl of DNA polymerase I (10 units/μl). The total reaction volume was made up to 40μl by the addition of water, mixed thoroughly and left to incubate at 16^{o}C for 90mins. Enzymes and unincorporated nucleotides were then removed using Quick Spin columns (Roche) containing G50 Sephadex beads as per the manufacture instructions. Labelled probes were eluted in TE pH7.5.

2.7.4 Quantification of Label Incorporation

Gridded nitrocellulose membranes were soaked briefly in water, followed by immersion in 20x SSC for 10 minutes and allowed to air dry. Labelled DNA was spotted onto these membranes in the following dilutions: 1:500, 1:1,000, 1:5,000 and 1;10,000 in TE and allowed to air dry. Onto the same membrane 20, 10, 2 and 1 pg of labelled lambda DNA standards were also spotted (Roche). Once all DNA spots had air dried, the DNA was crosslinked onto the membrane by exposure to 150mJ of UV irradiation.

The crosslinked membrane was immersed in buffer 1 (0.1M Tris-HCl pH7.5, 0.15M NaCl) for 5 minutes at RT with constant agitation, before it was blocked in buffer 1 supplemented with 3% BSA (w/v) for 60mins at 60^{o}C. After blocking the membrane was incubated at RT, with either strepavidin-alkaline phosphatase, and/or anti-digoxigenin-alkaline phosphatase (both diluted 1:1,000, Roche) as appropriate. It was then washed twice for 15mins at RT in buffer 1 before being equilibrated for 5mins in 0.1M Tris-HCl (pH 9.5). To develop the membrane, it was incubated in a sealed polythene bag with 5ml of 0.1M Tris-HCl (pH9.5) containing two drops from

each of bottles 1-3 from the alkaline phosphatase substrate kit IV (Vector Laboratories). The substrates in this colour reaction are 5-bromo-4-chloro-3-idolyl phosphate and nitroblue tetrazolium, which produce a blue reaction product. A complete colour reaction could be observed within a few hours and an estimate of the concentration of label incorporated into the DNA obtained from comparison with the standards.

2.7.5 Slide Preparation for FISH

Glass slides were stored in a dilute solution of HCl in ethanol for at least one hour prior to use. Slides were then dried and polished with muslin before use. MAA fixed cells (see above) were removed from storage at -20°C and centrifuged at 1000 g for 5 min. Fresh MAA fix was then added until the cell suspension reached a 'milky' appearance. One drop of this suspension was then dropped onto a horizontal microscope slide from a height of ~30 cm using a fine tipped pastette (to obtain good chromosomal spreads slides were preferably dropped when air humidity was ~50% and slides coated with a thin layer of moisture, usually achieved by breathing on the slides immediately before and after dropping). The quality of the spread was monitored by phase contrast microscopy. Slides were preferably stored for 2-6 days prior to hybridisation, however if slides were to be used the following day, they were artificially aged by baking at 60^{o}C for 1 hour prior to FISH.

2.7.6 Hybridisation of FISH Probes

Slides were treated with 100 μg/ml RNaseA in 2x SSC for 1 hr at 37°C, washed briefly in 2x SSC and dehydrated through an ethanol series (2 min each in 70%, 90% and 100% ethanol). Slides were then allowed to air dry before being heated in a 70°C oven for 5 min and then denatured in 70% formamide (v/v) in 2x SSC (pH 7.5) at 70°C for 90 seconds (except for preparations of 575 lymphoblastoid cells which were denatured for 50 seconds). Following denaturation, slides were immediately

plunged in 70% ethanol at 4°C for 2 min before further dehydration through 90% and 100% ethanol at RT.

Labelled probes were prepared by precipitation of ~150ng of labelled DNA with 5 μg salmon sperm DNA per slide and 10μg of human Cot1 DNA per probe. After the addition of 2 volumes of ethanol, probes were spun down under a vacuum before being resuspended in 10μl of FISH hybridisation buffer per slide. All chromosome paints used were commercial and therefore supplied in their own hybridisation buffer in a ready to use format (13μl paint per slide, Cambio). All probes were denatured at 70°C for 5 min and reannealed at 37°C for 15 min before being spotted onto coverslips and picked up by slides. Slides were sealed with rubber solution (TipTop) before incubation overnight in a covered tray in a 37°C water bath.

2.7.7 Washing and detection of FISH signal

After overnight hybridisation, rubber solution was removed from the coverslips and the slides washed 4 times in 2x SSC at 45°C for 3mins. The coverslips were allowed to fall off slides naturally during the first FISH wash. Slides were washed a further 4 times 0.1x SSC at 60°C for 3mins before being transferred to 4x SSC with 0.1% Tween 20 (v/v). Detection was carried out in a moist chamber pre-heated to 37°C. Biotin was detected with sequential layers of fluorochrome-conjugated avidin (FITC- or TR- avidin), biotinylated anti-avidin, and a further layer of fluorochrome-conjugated avidin. Digoxigenin was detected with sequential layers of Rhodamine (R)-conjugated anti-digoxigenin and TR-conjugated anti-sheep IgG. Detection reagents were diluted in 4xSSC supplemented with 5% Marvel milk powder (w/v) to the appropriate concentration (*Table 2.3*). Before use they were mixed thoroughly and centrifuged at 12,000g for 15mins to precipitate any clumps of antibody. Slides were blocked with milk supplemented 4xSSC for 5mins at RT before the first detection layer was applied to each slide. Slides were incubated with each antibody layer for 30-60mins in a moist chamber at 37°C. In between antibody layers and following the last antibody incubation slides were washed 3 times in 4xSSC with

0.1% Tween 20 (v/v) at 37°C for 2mins. All slides were mounted in 4,6-diamino-phenylindole (DAPI, ~1ng/ml) in Vectashield. Coverslips were sealed with rubber solution (PANG) and slides were stored in the dark at 4°C until imaged.

Table 2.3

Antibodies and Fluorochrome-Conjugates used for FISH

Antibody or fluorochrome-conjugate	Species	Supplier	Stock concentration (mg/ml)	Dilution
FITC-avidin	Goat	Vector labs	2.0	1:500
TR-avidin	Goat	Vector labs	2.0	1:500
Biotinylated anti-avidin	Goat	Vector labs	0.5	1:100
R-anti- digoxigenin	Sheep	Roche	0.2	1:20
TR-anti-sheep (IgG, H&L)	Rabbit	Vector labs	0.5	1:100

2.7.8 Capture of Fluorescent Images

After FISH, images were captured using either a Zeiss Axioplan II or Zeiss Axioplan fluorescence microscope, both using 100 watt mercury bulbs and equipped with a triple band-pass filter (Chroma # 83000). Grey scale images for each fluorochrome were collected with a cooled CCD camera depending on the model of microscope (Pentamax with a Kodak KAF 1400 sensor or Micromax with Kodak KAF 1400e sensor respectively, Princeton Instruments) using IPLAB software v.3.9 (Scanlytics, USA).

Slides were scanned in a methodical manner beginning at the bottom left hand corner, scanning to the right, and then moving upwards and scanning the next row of nuclei from right to left etc. For nuclear location analysis, (*Section 2.11.1*, below) sixty bin 2 images (using x63 oil immersion objective, Zeiss) were collected of consecutive nuclei that fulfilled the inclusion criteria and for interphase compaction analysis (*Section 2.11.2*, below), one hundred images were collected in a

similar manner. Criteria were that nuclei should be intact and contain visible signals. In addition for the purpose of running scripts, only single nuclei not touching any other were imaged.

2.7.9 Genomic Clones Used for FISH

The universal names and genomic locations of the BAC and PAC clones used to analyse cell line karyotypes and the nuclear position of the EDC are given in *Table 2.4*. 6 pairs of Fosmid clones were used to analyse the interphase compaction of the EDC and β-globin locus. The names and locations of the individual clones used are given in *Table 2.5* and the composition of each pair in *Table 2.6*.

Table 2.4

Names and Genomic Positions of BAC Clones Used

The universal names of each probe are given along with their positions. All positions are taken from the NCBI Build 36 (Ensembl release 46 August 2007). Note RP1-19K8 is actually a P1 derived artificial chromosome (PAC) but is generally referred to in this text as a BAC for convenience. All clones are located on human chromosome 1. The BACs were used to assay the karyotype of the EDC in cell lines and to assay the nuclear position of the EDC (see *Chapter 4*)

Name	Universal Name	Type	Start	End
Probe 1	RP1-19K8	PAC	151,357,640	151,409,748
Probe 2	RP11-216N14	BAC	151,954,673	152,104,838
Probe 3	RP11-172I6	BAC	154,346,199	154,350,727

Table 2.5

Names and Genomic Positions of Fosmids

The universal names of each Fosmid probe are given along with their positions. All positions are taken from the NCBI Build 36 (Ensembl release 46 August 2007). The Fosmids were used in pairs to assay nuclear compaction across the EDC (see *Chapter 4*)

Fosmid	Chromosome	Start	End
WI2-553L9	1	149,547,440	149,585,037
WI2-2080I19	1	150,097,276	150,139,393
WI2-535P8	1	150,216,505	150,253,863
WI2-885J7	1	150,722,348	150,761,412
WI2-1328A22	1	151,157,681	151,199,648
WI2-1606L15	1	151,418,410	151,457,452
WI2-3089H20	1	151,890,456	151,929,057
WI2-1690C15	1	152,430,849	152,468,880
WI2-528M6	11	4,961,240	4,999,789
WI2-2033J5	11	5,453,348	5,494,460

Table 2.6

Names of Fosmids Used in each Pair of Probes

The two Fosmid clones from *Table 2.5*, above, used in each pair referred to in *Chapter 4* are listed below. The distance in base pairs between the centres of the two Fosmids are also given. The EDC pairs are on human chromosome 1 and the Beta Globin pair on chromosome 11.

Pair	5' Fosmid	3' Fosmid	Distance
EDC Pair 1	WI2-553L9	WI2-2080I19	552,096
EDC Pair 2	WI2-535P8	WI2-885J7	506,696
EDC Pair 3	WI2-885J7	WI2-1328A22	436,785
EDC Pair 4	WI2-1606L15	WI2-3089H20	471,826
EDC Pair 5	WI2-3089H20	WI2-1690C15	540,108
Beta Globin Pair	WI2-528M6	WI2-2033J5	493,390

2.8 Preparation and Processing of Chromatin

2.8.1 Digestion Timecourse Analysis of Chromatin

Freshly prepared nuclei (*Section 2.6.5*, p74) were resuspended in NB-R to a DNA concentration of 0.2mg/ml. RNase A was added to 100µg/ml and 1ml aliquots of nuclei were brought to RT. 25 Units of Micrococcal Nuclease (MNase, Worthington) were added to each aliquot and after 1, 2, 4, 8, 16 and 32mins 100µl aliquots removed into an equal volume of genomic lysis buffer. DNA was then prepared from each aliquot (*Section 2.3.1*, p56).

1µg of DNA from each timepoint was then resolved on a 1.2% TBE agarose gel (*Section 2.3.5*, p57) which was stained with EtBr and scanned on a phosphoimager. Images of gels were analysed using the Aida program (Raytek) for 1D analysis. For repeat length analysis, appropriate DNA size markers (100bp and 1Kb ladders, Promega) were used to calculate the peak DNA size of each visible nucleosome fragment (ie mono-, di-, tri, etc). Linear regression of a graph of fragment number against DNA size was used to calculate the repeat length at each timepoint.

For digestion rate analysis, each lane was divided into two sections, high molecular weight and low molecular weight DNA. The border between the two was defined as half way between the penta- and hexanucleosomal fragment. The amount of DNA in each was calculated by measuring the total signal in each half of the gel. The percentage of high molecular weight DNA was then calculated.

2.8.2 Preparation of Soluble Chromatin for Sucrose Gradient Sedimentation

Large fragments of soluble chromatin were prepared by digesting freshly prepared nuclei with MNase and leaving them on ice overnight so that the chromatin was released from the nuclei. Nuclei (prepared as, *Section 2.6.5*, p74) were resuspended in NB-R buffer to a DNA concentration of 1mg/ml. Three 1ml aliquots were then equilibrated at RT and RNase A was added to 100µg/ml. These three aliquots were then digested with 2, 4 and 6 Units of MNase respectively for 9mins at RT. The digests were stopped by the addition of EDTA to a concentration of 10mM and the nuclei spun down gently before they were resuspended in cold $TEEP_{20}$ buffer. They were then left overnight on ice and the next morning the nuclei spun down and the supernatant containing soluble chromatin removed.

2.8.3 Sucrose Gradient Fractionation of Chromatin

Soluble chromatin was fractionated on sucrose gradients prepared with $TEEP_{80}$ buffer. Isokinetic 6-40% sucrose gradients were prepared in SW41 tubes with 0.5cm left at the top for the chromatin to be loaded. 450µl of chromatin was loaded per gradient and centrifuged (3.5hrs, 41,000rpm at $4^{o}C$) in a SW41 rotor (Sorvall). 500µl fractions were collected by upwards displacement and then either processed to purify DNA (*Section 2.3.1*, p56) or TCA added to precipitate proteins (*Section 2.5.2*, p69). Material from the gradient was passed through a UV monitor and their absorbance at a wavelength of 254nm quantified. From the average A_{254} of each fraction the DNA concentration of each fraction could be calculated (equating A_{254} with A_{260}, *Section 2.3.4*, p57).

2.8.4 Analysis of Bulk Chromatin

To analyse bulk chromatin, 1.5µg of DNA from sucrose gradient fractions was run on TPE agarose gels overnight (*Section 2.3.5*, p57). The next day these gels were stained with EtBr and scanned on a phosphorimager. Appropriate DNA size markers (1Kb, Promega and 2.5Kb, Biorad) were used to calculate the DNA size of the intensity peak in Aida.

2.8.5 Purification of Open Chromatin

To purify open chromatin, 3µg of DNA from a single high molecular weight chromatin fraction was resolved on a pulsed field gel (*Section 2.3.6*, p58). After staining the DNA size of the intensity peak was estimated from the molecular weight markers under ultra-violet illumination. A band was cut from the gel corresponding to DNA with a molecular weight of roughly twice this. The DNA was then extracted from the agarose as described in *Section 2.3.7*, p59.

2.8.6 Preparation of Chromatin for ChIP

To prepare chromatin for ChIP, freshly prepared nuclei were resuspended to a DNA concentration of 0.5mg/ml in NB-ChIP. To prevent the deacetylation of histones 5mM Na Butyrate was added to buffers used in the isolation of nuclei. 0.5ml aliquots of these nuclei were brought to RT and 100µg/ml of RNase A added. 50 U of MNase was added and chromatin digested for 10mins at RT with gentle mixing. The reaction was stopped by the addition of 0.5ml of NB-Stop which brought the buffer composition to roughly (150mM NaCl, 20mM Tris HCl (pH8), 10mM EDTA, 0.1% Triton X-100 (v/v), 5.5% Sucrose (w/v)). Nuclei were then left on ice overnight. The next morning nuclei were spun down and the supernatant containing soluble chromatin removed.

2.8.7 Immunoprecipitation of Native Chromatin

Soluble chromatin was first precleared. This was done by adding 50µl of a 50% slurry of protein A sepharose beads (Roche) to the chromatin solution and incubating the mixture on a wheel at 4^{o}C for 2hrs. Before use the beads were resuspended in TE and blocked with 10µg/µl of BSA. Precleared chromatin was separated from the beads by spinning the mixture for 2mins at 2K and 4^{o}C in microcentrifuge and removing the supernatant to a fresh microcentrifuge tube. 5 µg (for anti-H4K9Ac and anti-H4K16Ac) or 10µg (for anti-H3K9diMe and anti-H3K27diMe) of antibody was then added to the chromatin and it was incubated overnight at 4^{o}C on a rotating wheel. For mock precipitations 5µl of Rabbit Serum (Scottish Antibody Production Unit) was added instead of the antibody. For precipitations with anti-H4K16Ac the chromatin was diluted 4x with a 1:1 mix of NB-ChIP and NB-Stop before preclearing and for precipitations with anti-H3K9diMe and H3K27diMe the chromatin was diluted 2x in the same manner.

The following morning 50µl of blocked protein A beads were added to the chromatin and it was incubated for 2hrs on a wheel at 4^{o}C to bind the antibody-chromatin complexes to the beads. The beads were then spun down, the supernatant removed and the beads washed with a series of wash solutions for 5mins at 4^{o}C on a wheel (for wash solutions see *Table 2.7*). After the third wash the beads were washed twice in the same manner with TE. Then 250µl of ChIP elution buffer was incubated with the beads for 15mins at 37^{o}C with shaking to elute antibodies and chromatin. The beads were then spun down, the supernatant removed and a further 150µl elution performed in the same manner. The eluted chromatin could then be used to make protein (as *Section 2.5.2*, p69) or DNA (as *Section 2.3.1*, p56). These conditions were empirically determined to specifically precipitate nucleosomes using an antibody and not rabbit serum or protein A beads alone. In addition the specificity of each antibody was checked by western blot of nuclear extracts of 293 and HaCaT cells (see *Section 2.5.5*, p70 and *Figure 2.2*)

Table 2.7

Washes for ChIP

The antibodies and washed conditions used are given below, for details of the antibodies used see *Table 2.2*, p70. All washes were supplemented with 5mM Na Butyrate, 250µM PMSF and Complete Proteinase Inhibitors (Roche)

Antibodies	**Anti-H3K9Ac**	**Anti-H4K16Ac, H3K9diMe, H3K27diMe**
Wash 1	25mM Tris pH8, 2mM EDTA, 150mM NaCl, 1% Triton X-100 (v/v), 0.1% SDS (w/v)	25mM Tris pH8, 2mM EDTA, 150mM NaCl, 0.05% Triton X-100 (v/v)
Wash 2	25mM Tris pH8, 2mM EDTA, 500mM NaCl, 1% Triton X-100 (v/v), 0.1% SDS (w/v)	25mM Tris pH8, 2mM EDTA, 200mM NaCl, 0.1% Triton X-100 (v/v)
Wash 3	10mM Tris pH8, 1mM EDTA, 250mM LiCl, 1% Na Dexoycholate (w/v), 1% Igepal-360 (v/v)	25mM Tris pH8, 2mM EDTA, 150mM LiCl, 0.1% Na Dexoycholate (w/v)

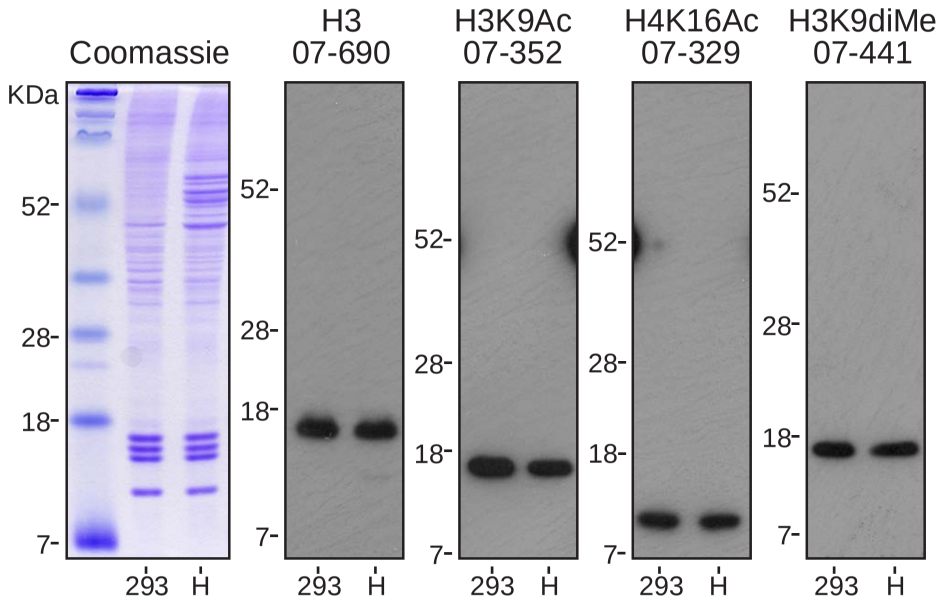


Figure 2.2 Specificity of Antibodies used for ChIP

Nuclear extracts of both HaCaT (H) and 293 cells were blotted with antibodies against histone modifications or stained with coomassie. The names of the antibodies and their catalogue numbers are given above each blot (All antibodies supplied by Upstate). Gels were run on different days and hence for varying lengths of time. Therefore, the different histone H3 and H4 bands do not always line up despite being the same molecular weights.

2.8.9 SYBR Gold Staining of ChIP Gels

DNA was purified from chromatin eluted from ChIP and 5µl (10%) resolved on 1.2% TBE agarose gels (see *Section 2.3.5*, p57). The gels were then stained in TBE with SYBR Gold (Invitrogen) and briefly destained in TBE. They were then scanned on a phosphorimager using a 473nm laser and a 510nm long pass filter (LPB filter Fuji FLA-5100 phosphorimager). Experiments using serial dilutions of salmon sperm DNA had shown that ~50pg could be detected on the gel (data not shown). The amount of DNA present in precipitated samples was estimated by densitometry using a serial dilution of DNA purified from input chromatin.

2.8.10 Crosslinking ChIP

Crosslinking ChIP was performed with the help of C. Kress. Two 9cm dishes of confluent HaCaT cells had 2ml of Crosslinking Solution added to 20ml of media. They were then incubated for 10mins at RT with gentle mixing before 20ml of PBS supplemented with 0.125M Glycine was added to quench the crosslinking. The cells were then washed with PBS and then scraped from the plates in ice cold Chromatin Wash 1 (Chr-W1 solution). The crosslinked cells were then washed for 10mins on ice with shaking before being spun down and washed again in Chr-W2. Finally they were spun down again and washed in Chr-W3. The pellet was then resuspended in X-ChIP sonication buffer and sonicated for 15mins in 30s on/off cycles on the high setting using a Bioruptor sonicator (Cosmo Bio). The chromatin was then quantified by its DNA concentration measured by its A_{260} and the size of fragments checked from an aliquot by reversing crosslinks overnight at 65^{o}C and purifying the DNA. A range of fragments of average size ~500bp was observed. The rest of the chromatin was supplemented with 5% glycerol (v/v) and stored at -70^{o}C for precipitation later.

To immunoprecipitate the chromatin, 50µg of chromatin was supplemented with 1% Triton X-100 (v/v) and precleared as for Native Chromatin (see above). Immunoprecipitation with anti-H3K9Ac, washing and elution of chromatin were all

carried out as for Native ChIP. However, prior to cleaning up the DNA for analysis crosslinks had to be reversed by incubating the chromatin overnight at 65^{o}C.

2.8.11 Preparation of Mononucleosomes

To prepare mononucleosomes, freshly prepared nuclei were resuspended to a DNA concentration of 0.5mg/ml in NB-ChIP. 1ml aliquots of these nuclei were brought to RT and 100µg/ml of RNase A added. 200U of MNase was added and chromatin digested at RT with gentle mixing. Aliquots of nuclei were removed to genomic lysis buffer after 4, 6, 8, 10 and 12mins to stop the reaction. DNA was then purified from these aliquots and analysed on an agarose gel. The 12min timepoint contained no high molecular weight DNA and mono-, di-, tri- and tetranucleosomes and was therefore suitable for analysis (see *Chapter 5*). DNA from this fraction was therefore resolved on a 4% NuSieve GTG agarose gel stained with EtBr and visualised under ultraviolet illumination. The area of the gel containing mono-nucleosomes was cut out with a clean razor blade and the DNA extracted from the gel as *Section 2.3.7* p59.

2.9 Microarrays

When referring to microarrays I have adopted the convention that the DNA molecules attached to the slides are referred to as probes and the labelled DNA hybridised onto the slides is the target DNA (Phimister, 1999).

2.9.1 Whole Genome Expression Arrays

Whole genome expression arrays were acquired from B. Ylstra (VU University Medical Center, Amsterdam). These were composed of oligonucleotide probes from the Human Release 2.0 oligonucleotide library, which contains 60mer oligonucleotides representing 28830 unique genes as designed by Compugen (San Jose, CA, USA, van den Ijssel et al., 2005). They were printed on CodeLink™ slides

(Amersham Biosciences) which use a proprietary surface chemistry to covalently link oligonucleotides to a polymer matrix which also serves to lift the probe above the glass surface, preventing steric hinderance (Southern et al., 1999).

2.9.2 Custom EDC Oligonucleotide Arrays

Customised oligonucleotide arrays were made from 5000 oligonucleotide probes synthesised by Illumina in collaboration with Invitrogen. The probes are 60-mer oligonucleotides which were designed to our specifications by P. Rigault (Centre de Génomique de Québec, Québec). 2145 oligonucleotides were designed to cover the EDC with a resolution of one oligonucleotide per ~792bp. This resolution varies because extra oligonucleotides were often designed in genic and promoter regions and because the presence of repetitive sequences prevented the design of oligonucleotides in some regions. 814 oligonucleotides are tiled at a slightly lower resolution (1 every ~883bp) outside of the EDC for ~360Kb. The rest of the oligonucleotides are spread around the genome, although there are higher concentrations in the region around the EDC, the cytokeratin clusters and some regions that may be of future interest in the laboratory. Probes were checked using two criteria against the human genome, firstly that they did not contain a 55bp stretch of sequence that was 90% similar to another region of the human genome and secondly that they did not contain a 30bp stretch that was identical to another part of the human genome. Probes were also designed so that they all had roughly the same melting temperature.

The arrays themselves were printed by the Liverpool Microarray Facility with the help of M. Hughes (University of Liverpool) onto Codelink slides. The probes were synthesised in a random order in 96 well plates so that they were printed in a random arrangement on the slides. They were also synthesised with a 5' reactive amine to enable them to be covalently crosslinked to the slide surface. Oligos were printed at a concentration of 20μM in 1x Nexterion spot buffer (Schott) containing 0.005% (v/v) Triton X-100.

2.9.3 Hybridisation of Whole Genome Expression Arrays

Labelled Cy3 and Cy5 target cDNA samples (prepared as *Section 2.4.6*, p68) were mixed with 10µg of salmon sperm DNA and 500µg of yeast tRNA before being ethanol precipitated. The pellet was then resuspended in Microarray Hybridisation Buffer for 1hr at 37^{o}C before being denatured at 95^{o}C for 5mins and cooled on ice for 5mins. They were then applied to the arrays under an unsealed, ethanol cleaned coverslip. Arrays were hybridised overnight at 42^{o}C on an enamel tray in a water bath and in a sealed chamber humidified with tissues soaked in 5xSSC.

After hybridisation the slides were removed to a large volume of wash solution and the coverslips allowed to fall off naturally during the first wash. Washes were twice with 2xSSC, 0.1% SDS (w/v) at 42 ^{o}C, twice with 0.2x SSC at RT and twice with 0.1xSSC at RT (all for 5mins each). After the final wash the slides were briefly plunged into 0.01xSSC at RT before being spun dry and immediately scanned. All cDNA samples were cohybridised with the Stratagene Universal Human Reference RNA (Stratagene) as a reference samples (it consists of a mix of RNAs extracted from 10 different human cell lines).

2.9.4 Hybridisation of EDC Arrays

Labelled target DNA samples (prepared as *Section 2.3.9*, p61) were mixed with 10µg of Human Cot1 DNA and 1.3mg of Yeast tRNA before being ethanol precipitated. They were then resuspended in Microarray Hybridisation Buffer as for Whole Genome Expression arrays and after denaturation and cooling hybridised in the same manner except that hybridisation was performed at 45^{o}C for ~36hrs. I determined that the longer and more stringent hybridisation produced better quality array data (data not shown). Washing was also performed as for whole genome expression arrays except that the first wash was at 45^{o}C.

2.9.5 Scanning of Hybridised Microarrays

Arrays were scanned on a ProScan Array HT scanner (Perkin Elmer). Cy3 and Cy5 channels were scanned separately using automatic intensity calibration settings built into the system. For Cy3 a 543nm laser and 570nm band pass filter were used and for Cy5 a 633nm laser and 670nm band pass filter were used. Briefly a representative area of the array was used to vary the Photomultiplier Gain (PMT Gain) of the scanner such that the most intense spots in that array were around ~90% saturated. This ensured that each channel was scanned to give the greatest dynamic range of data. In some cases this was then manually adjusted to give a scan with lower levels of background noise. In addition, some arrays (see *Chapter 5*) were scanned at the same PMT gain to allow direct comparisons. All arrays were scanned using a laser power of 90% and at a 10μm resolution. Scans were collected and visualised using the ScanArray Express Software (Perkin Elmer) and each channel saved separately as 16-bit tagged-image file format (*tiff*) files.

2.9.6 Processing of Individual Array Images

Individual microarray scans were analysed using the Bluefuse v3.2 software package (Bluegnome). Images of each channel of the hybridised arrays were loaded into the software together with a Genepix Array List (*gal*) file giving a the identifier (ID) number of the oligonucleotide probe printed at each spot (for Whole genome arrays it was provided by B. Ylstra, and for the EDC arrays it was generated by the University of Liverpool during printing). A further file giving the genomic location (NCBI build 36) of each probe, as defined by its ID number, was also used (*cfg* file, see Bluefuse documentation, for Whole genome arrays it was provided by B. Ylstra and for EDC arrays it was generated from data provided by P. Rigault after probe design).

Bluefuse was used to extract spot intensities and process background signal for each array. It was also used for normalisation of array data and to calculate average values for each probe on the array if it was printed multiple times. For

whole genome expression arrays, each probe was printed once and so no averaging was used. The two channels were normalised to each other using a block lowess strategy because it is capable of correcting both for intensity dependent artefacts in $\log_2$ ratios (deviation from a mean ratio $\log_2$ ratio of 0 for low and high intensity spots) and for systemic spatial variation in the array caused by, for example, slight imperfections in the slide surface or local variations in hybridisation conditions (Bluefuse settings for lowess, f=1/4, iter=3). These arrays contained 625 spots per printed block and spots were arranged at random on the array, satisfying the criteria for this normalisation strategy (Quackenbush, 2002). Lowess normalisation, is based on the fit of a locally weighted scatter plot model and was first used by the laboratory of T. Speed (Yang et al., 2002)

On the other hand, although the EDC arrays were printed in blocks of 462 spots, each of these block contained 4 replicates of each probe. This means the actual number of probes per block was 115, which is not sufficiently high for a block lowess normalisation strategy. Therefore a global lowess strategy was used instead. To generate average values for the 4 printed replicates of each probe I used the 'fusion' setting for replicate combination in BlueFuse. This is a proprietary method which makes use of confidence values generated for each spot in Bluefuse (based on a variety of parameters such as intensity). When combined by this method, low confidence replicates have a lesser influence on the final values than those with greater confidence values. The results of individual array analyses were exported from Bluefuse as tab delineated Excel files.

2.10 Analysis of Data

2.10.1 Defining the EDC in the Genome

All genomic positional data was taken from NCBI build 36 and generally viewed in the UCSC genome browser (March 2006 build). All gene positions and intron structures were taken from the Refseq annotation in the same genome build.

The EDC was defined as beginning halfway between its most 5' gene (*S100A10*, Refseq: NM_002966.2) and the nearest gene lying outside the EDC (*THEM4*, NM_053055.3). Similarly it defined as ending halfway between the *S100A1* gene (NM_006271.1) and *SNAPAP* (NM_012437.3). This meant the co-ordinates of the EDC were human chromosome 1: 150185374-151884481.

The various gene clusters of the EDC were defined in a similar way, by identifying their outermost genes and defining the boundaries of the clusters as halfway between them and their nearest neighbouring gene in the Refseq annotation (*Table 2.8*). The information on probe position for the Whole Genome Expression Arrays was taken from the July 2003 NCBI 34 assembly and were manually mapped onto the NCBI 36 assembly around the EDC.

The positions of repeats were taken from the UCSC genome browser Repeatmasker annotation (March 2006 build). The table browser combined with filters was used to download the positions of repeats of either the SINE or LINE class in the region. Similarly the annotation of GC content was taken from the UCSC browser annotation which is in 5bp windows.

Table 2.8

Genomic Co-ordinates of Gene Clusters of the EDC in NCBI build 36

The names and Refseq identifiers of each cluster's bounding gene are given along with the resulting genome co-ordinates for each cluster. All co-ordinates are for human chromosome 1 in NCBI build 36. The *S100A*-1 cluster is the centromeric cluster consisting of the *S100A10* and *S100A11* genes whereas *S100A*-2 is the telomeric cluster containing the rest of the *S100A* genes.

Cluster	Bounding Genes	Refseq Ids	Start	End
S100A-1	S100A10 S100A11	NM_002966.2 NM_005620.1	150185374	150299689.5
Fused	THHL1 CRNN	NM_001008536.1 NM_016190.1	150299689.5	150701648
LCE	LCE5A LCE1A	NM_178438.4 NM_178348.1	150701648	151100252
SPRR	SPRR4 SPRR2G	NM_173080.1 NM_001014291.2	151180369	151415887
S100A-2	S100A9 S100A1	NM_006271.1 NM_002965.3	151592300	151884481

2.10.2 General Statistical and Quantitative Analysis

General quantitative analysis was carried out using the Excel spreadsheet program (Microsoft). Data was transferred from other programs and analysed using default settings within Excel. Statistical analysis was either carried out using Excel or the R open source statistical language (mainly version 2.4.1). All graphs were generated in one of these two programs and modified using the Freehand MX vector graphics package (Macromedia). All boxplots are drawn using the default settings within R, where the median is represented by a heavy line, and the boxed area encompasses the range from the 25^{th} to the 75^{th} percentile of the data's distribution. The whiskers of the plot extend outwards from these percentiles to a point defined as 1.5 times the

distance between these two percentiles (*ie* 1.5 times the distance covered by the box). Outliers of this range are then indicated by circles.

In general, I have not assumed normality in my statistical analysis and made use of non-parametric statistical tests. Firstly, I have often compared a value of interest to a distribution of randomly sampled control values. The p-value in such a case is simply the percentage of the control distribution that has a value greater than or less than the tested value as appropriate. Secondly, I have used the Kolmogorov-Smirnov test (KS-test) to determine the probability that two groups of values are sampled from the same distribution. In general, non-parametric tests are thought to be less powerful than parametric statistics such as the students t-test. However, it is clear that most of the data examined in this study is not normally distributed, for example the distribution of d^2 values observed when measuring interphase compaction is known to conform to a Rayleigh Distribution (van den Engh et al., 1992).

I have also used non-parametric measurements of correlation where appropriate. However, when calculating the correlation between two replicates of a particular experiment, we expect the values to be normally distributed around the 'real' value, therefore I have used the Pearson's correlation which assumes normality. In other cases we cannot assume this, so I have used the Kendall's correlation which does not assume normality, or a linear relationship between the two values (Kendall, 1938). Spearman's correlation could also be used but, the statistic derived by Kendall's method generally has a simpler interpretation (Wilkie, 1980).

2.10.3 Defining Gene Promoters

To define gene promoters, gene annotations for chromosome 1 were downloaded from the UCSC browser using the table browser function (Refseq gene annotation). These consisted of a Refseq gene ID, the strand of transcription (+ or –), and a start and end coordinate for the gene (both with regard to the genomic sequence). Using a script in the R program, I converted these co-ordinates to transcriptional start sites. For genes transcribed on the + strand, the transcriptional start was the start

coordinate of the gene and for those on the – strand, the end coordinate. A further script was then used to identify the id numbers of oligonucleotides which lay within a defined distance of these transcriptional start sites (generally I defined the window as +/- 1000bp from a transcriptional start). The subselection of groups of promoters, was made using the gene cluster boundaries defined above.

2.11 Analysis of Fluorescence Microscopy Images

2.11.1 Analysis of Nuclear Position

The nuclear position of the EDC was measured relative to its parent chromosome 1 territory. FISH was performed as described in *Section 2.7* and 60 images taken in a systematic manner from each hybridised slide. In each image I then calculated the nuclear position of each BAC probe signal based on its distance from the nearest chromosome 1 territory edge using the IPLab (v3.9.5) software and a script initially developed by P. Perry and N. Mahy (Mahy et al., 2002b). In my case, the chromosome territory was always labelled in the FITC channel and the BAC probes in the TxRd channel, so the analysis is described in these terms. Briefly, background was removed from the FITC and TxRd signals by automatic calculation of the level of background within the nucleus. The TxRd signals were then automatically segmented and the signals of interest indicated by the user. The coordinates of the weighted centroid of the segment of interest were then calculated and stored.

The FITC signal from the chromosome territory was then segmented with user input, and the territory of interest within the nucleus defined by the user (in my case the closest territory to the signal of interest). Each pixel in the image was then defined as either containing chromosome territory signal or not. Then a segmentation disc was dilated out, one pixel at a time, from the TxRd signal centroid. This was done until the expanding disc met the edge of the chromosome territory segmentation region, *ie* nearest chromosome territory edge. The distance of the

probe from the territory edge was then defined as the radius of the expanded signal segmentation region. For probe positions internal to the territory this distance was annotated as +ve and those outside as –ve.

Every probe signal in each nucleus was handled separately in this manner and the data exported to Excel. In Excel, the raw distances which were in pixels were converted to microns using a scaling factor (calculated from the parameters of the microscope and lens used, 1 pixel = 0.134μm, P. Perry). The distribution of distances for each probe and cell line combination could then be analysed.

2.11.2 Analysis of Interphase Chromatin Compaction

Interphase chromatin compaction was analysed using an adaptation of a script originally written by P. Perry (Chambeyron and Bickmore, 2004). With the help of P. Perry, I modified the script to facilitate more rapid analysis of large numbers of pictures. FISH was performed as described in *Section 2.7* and 100 images taken in a systematic manner from each hybridised slide. Images for analysis were composed of 3 channels, the DAPI staining of nuclear DNA and two genomic probes which were labelled in FITC and TxRd (*Figure 2.3*). Firstly, the area of the nucleus was calculated from an automatic segmentation of the DAPI signal. Next the user identified pairs of genomic probes in the nucleus (ie pairs of FITC, green, and TxRd, red, signals). After automatic background reduction, the computer then segmented both the green and red channels. In each user defined region, the coordinates of weighted centroid of both the red and green channel were stored. To calculate the distance between the centre of each probe signal, the distance between these two coordinates was calculated by basic trigonometry: ie where distance = *z*, and the coordinates of each centroid given by *x* and *y*:

$$z^2 = (x_{green} - x_{red})^2 + (y_{green} - y_{red})^2$$

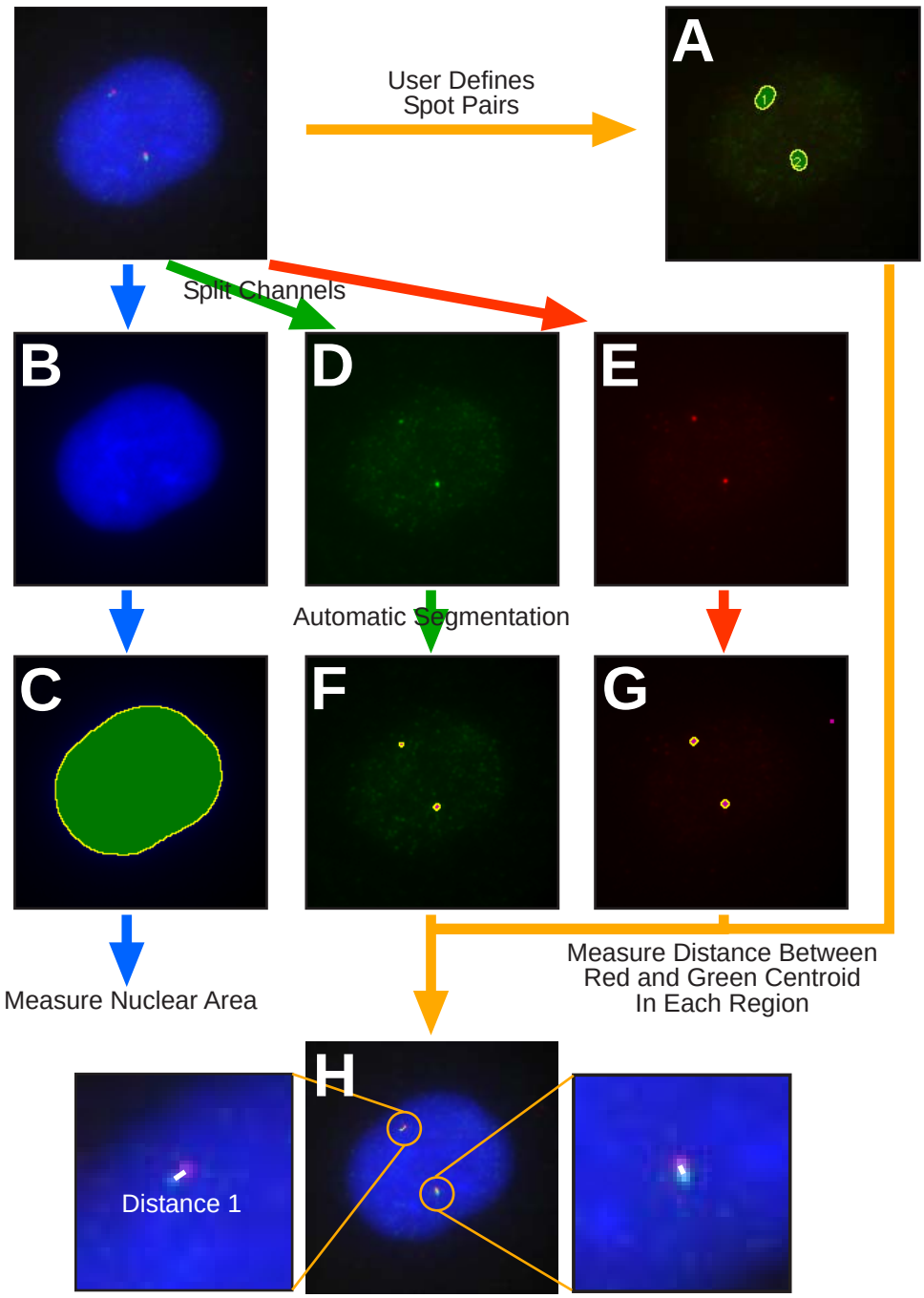


Figure 2.3 Analysis of Chromatin Compaction by the Distance Between Two Spots

To measure distances, the user first defines areas of the nucleus contatining a pair of red and green spots (A). The computer then splits the colour channels, calculating the nuclear area from the blue channel by segmentation (DAPI signal, B+C). The red and green channels (D+E respectively) are then segmented to identify spots (F+G). The centroid of each identified spot is then calculated and its coordinates stored. In each user defined region, the coordinates of its red and green spots are then used to calculate the distance between the spots by basic trigonometry (H).

Raw distances in pixels were then exported into Excel and converted into microns using the scaling factor of 1 pixel = 0.134μm (as above). There is a relationship between the distance observed to separate the probes in the interphase nucleus and the genomic distance separating the two clones used as probes (at least for genomic separations <2Mb, (van den Engh et al., 1992)). Where the mean distance separating a pair of probes in a sample of nuclei = d, d^2 is linearly related to the genomic distance separating the two probes. To compare the compaction at different regions of the EDC, I therefore adjusted the d^2 value by the genomic separation of each probe pair. I also calculated the normalised standard error in the mean to aid statistical comparison.

To calculate the area of the nuclei, the radius in pixels was first derived from the area in pixels outputted by the script. This was then converted to μm using the same scaling factor as above and the area in μm^2 calculated assuming a circular nucleus.

2.12 Analysis of Microarray Data

2.12.1 Dye Swap Analysis of Replicate Hybridisations

To combine, two normalised replicate hybridisations, I simply took the mean of their $\log_2$ ratios (and for more than two hybridisations, the mean of all was taken). However, first I elected to remove probes from the analysis using information from dye swapped replicate hybridisations using an adaptation of a published method (Quackenbush, 2002). Dye swapped controls for 2 colour microarray analysis mean that the hybridisation is replicated in alternating dye orientations (*ie* for one replicate the hybridisation is performed with Cy3 labelled sample and Cy5 labelled reference and for the other replicate with Cy5 labelled sample and Cy3 labelled reference). This type of control was originally performed to correct for the differential

incorporation of the two dyes into the labelled sample (Quackenbush, 2002). Some reports have suggested that lowess normalisation automatically corrects for this bias, however, I have observed probes which consistently behave in an inconsistent manner across the dye swap (data not shown).

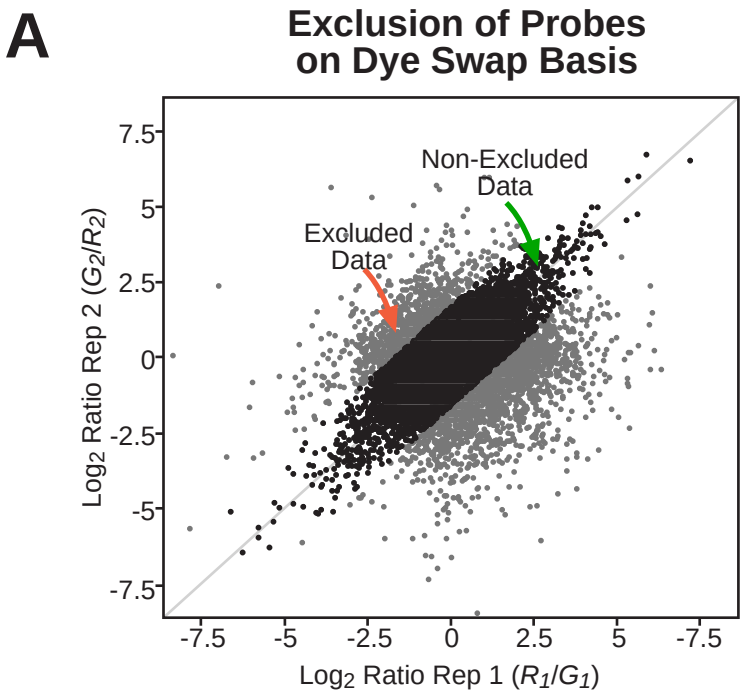
Therefore I performed replicate filtering using the dye swapped replicates to remove probes from the analysis which did not behave consistently across the dye swap. This was performed using an adaptation of a previously published method (Quackenbush, 2002). If we assume that the ratio of a particular probe, *T* is the result of dividing its intensity in one channel by the intensity of the other channel (Red, *R* and Green, *G*, respectively), then in theory:

For replicate 1: $T_1 = R_1/G_1$ should equal (in a dye swapped replicate): $T_2=G_2/R_2$

Therefore: $T_1/T_2 = 1$ or $log_2(T_1/T_2) = 0$

However, in practice this will not be the case, so we can define the ratio, T_1/T_2, as a dye factor, *D*. The objective of this filtering is identify the probes for which the log_2 *D* value deviates highly from zero and to exclude them from the analysis because they do not behave consistently across the dye swap.

Therefore, *D* is calculated for every probe on the array from their normalised ratios in two replicates. Then the distribution of log_2 *D* values for all probes is examined and its median (which will be close to 0), 25th and 75th percentiles are calculated. The interquartile distance is then defined as the distance from the 25th to the 75th percentile and any oligonucleotide whose $\log_2$ *D* value lies more than 2 times the interquartile distance from the median of the distribution was excluded from analysis (for a graphical representation, see *Figure 2.4A*). This is analogous to the method described by J. Quackenbush except that I use the median and a threshold based on the interquartile distance, whereas he describes the use of the mean and standard deviation (Quackenbush, 2002). Therefore, my adaptation of the method does not assume a normal distribution of log_2 *D*, but in the case of normal distribution, the methods will be essentially identical. The similarity of the methods can be seen by comparing *Figure 2.4A* to *Figure 3* in Quackenbush, 2002.



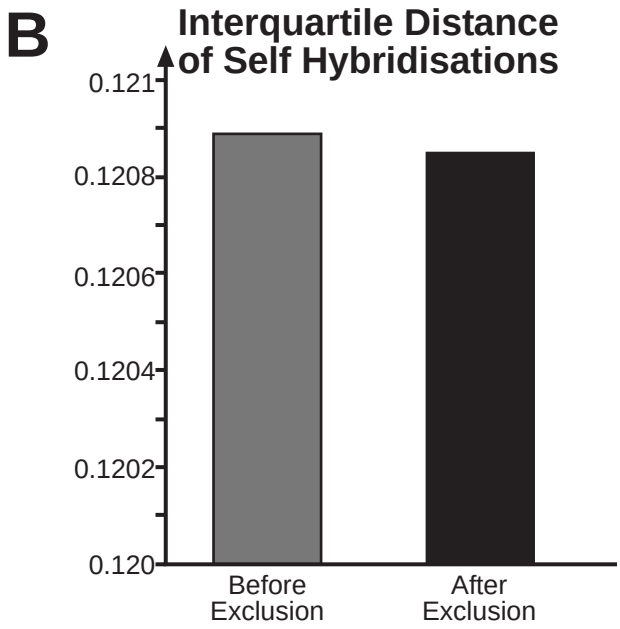


Figure 2.4 Excluding Probes from Microarray Analysis using Dye Swap Analysis of Replicates

A. Scatter plot showing data before and after exclusion, the $\log_2$ ratios from each replicate are plotted against each other. The dye factor, *D*, is the ratio of the two replicates. Points whose dye factor deviates greatly from the median *D* of 0 (grey line) are excluded from further analysis (grey points). The rest of the points (black) are kept.

B. Plot of the Interquartile distances of self hybridisation datasets. In theory, when a DNA sample is hybridised against itself the $\log_2$ ratio of every probe should be zero. However, this is not seen and instead ratios are distributed around 0. Therefore, the degree of scatter from zero, is a measure of the error in the experiment. The interquartile distance is the distance from the 25th to the 75th percentile and can quantify this error. After exclusion on a dye swap basis as in A., the interquarile distance, and therefore, experimental error decrease.

In the case of more than two replicates, the analysis was only changed in that the dye factor, *D*, was calculated from the mean ratio of a probe in each dye orientation (*ie* if the ratio of interest is *Sample/Reference* then T_1 is the mean $Sample_R/Reference_G$ ratio and T_2 is the mean $Sample_G/Reference_R$ ratio). In the analysis of gene expression in *Chapter 3*, the percentage of spots excluded per experiment was ~10% whereas in chromatin analysis using the EDC arrays, generally <5% of spots were excluded.

2.12.2 The Use of Self-Self Hybridisations to Determine Experimental Error

All microarray experiments have a degree of technical error and it is possible to gains some estimation of this using self hybridisations. In theory, if a single DNA sample is labelled with both dyes and hybridised against itself, the ratio of every probe on the array should be 1 (or a $\log_2$ ratio of 0). However, in practice what we observe is that the $\log_2$ values for all probes are scattered around a median of zero. If the error was systematic we would expect that two replicates of the self hybridisation would be correlated. When I calculate the correlation of two biological replicates of a self hybridisation, we find that there is in fact almost no correlation between their $\log_2$ ratios (Pearson's $R^2 = 0.079$), suggesting that the error is random.

The degree of error can be measured by quantifying the magnitude of the scatter in the distribution, for example by measuring the distance between the 25^{th} and 75^{th} percentiles (interquartile distance). When we do this, we find that the dye swap analysis described above reduces experimental error measured by self hybridisation (*Figure 2.4B*), presumably because the more unreliable data points are removed.

When analysing microarray experiments, values of interest are usually characterised by their degree of deviation from a $\log_2$ ratio of zero. For example, in expression analysis, deviation from this value indicates the gene represented by a particular probe is differentially expressed between two samples. However, as we have seen, $\log_2$ ratios can deviate from zero purely through experimental error. We can assign a p-value representing, for example, the probability that a particular probe

represents an upregulation of a gene by counting the percentage of probes in the self data set that are both greater than the self dataset median and greater than or equal to the $\log_2$ ratio of the probe of interest in the experiment.

Similarly we can analyse, groups of probes using the same idea. The deviation of a group of probes can be measured by their median $\log_2$ ratio. To assign p-values to them we simply take random samples of values from the control dataset of a size equal to the number of probes we are interested in. If we repeat this many times (1,000-1,000,000) and calculate the sample median each time, we build up a distribution of medians which measures the experimental error for a group of probe. P-values can then be assigned to experimentally measured medians in the same manner as for a single probe.

In general in this text, I have only considered a value to be above experimental error if its p-value, measured as described above, is vanishingly small. This implies that the observed $\log_2$ ratio had a very low probability of being observed through experimental error.

2.12.3 Sliding Window Analysis

When examining chromatin structure across a large region of the genome such as the EDC, it is often impractical to consider every probe individually. By examining mean $\log_2$ values from different parts of the locus we are able to gain a better overall understanding of how chromatin is structured across the region. By doing this we are also less likely to base hypotheses on the values of single probes, which are inherently more erroneous than the mean of a group of probes.

One method of looking at the trend in the data is to use a running mean, where a mean value for every datapoint is generated from groups of adjacent probes, for example groups of five (*Figure 2.5A*). However, this method has one problem, microarray probes on the EDC arrays are not evenly spaced across the region. Therefore, in some regions, the mean of 5 probes could represent ~1Kb whereas in others it could represent ~5-10Kb. Therefore, I generally calculated running means

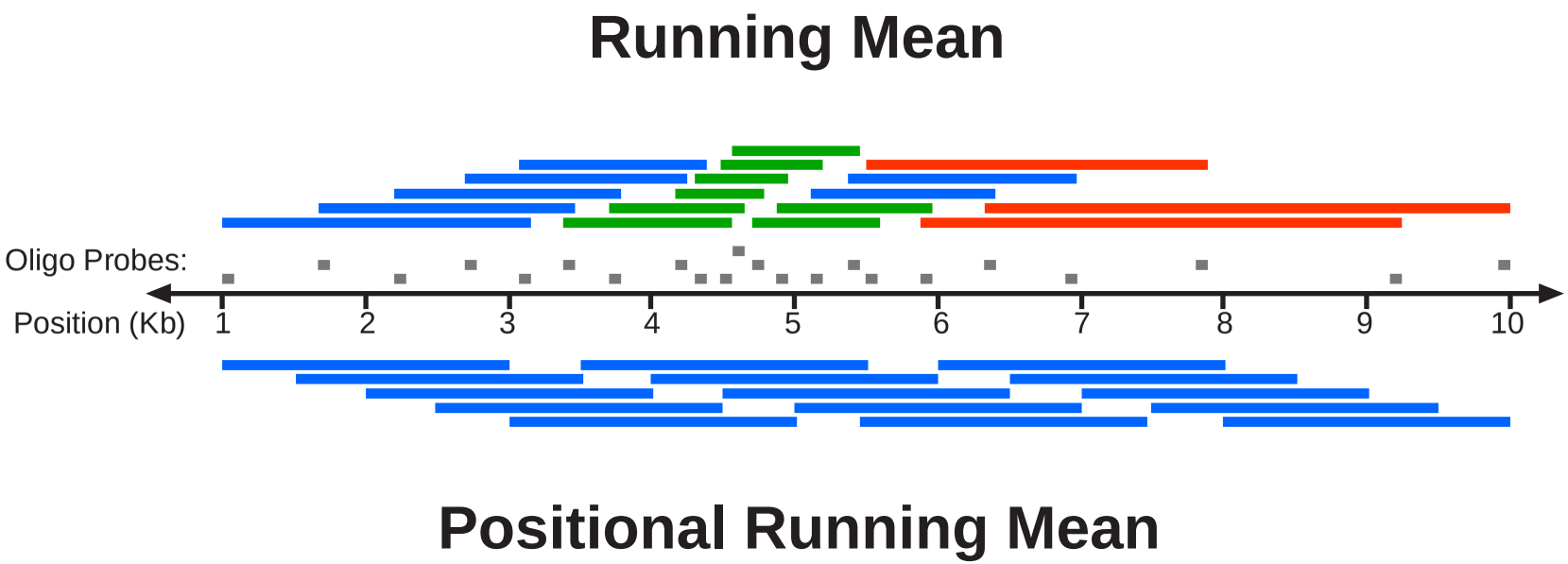


Figure 2.5 Comparison of Sliding Windows Based on Number of Oligos and Position

A running mean is generally calculated by passing a sliding window over every group of probes across a region (in this case 5 probes). However, although most windows will be of an average size (here ~2Kb, blue windows), the varying density of oligos on a tiling path oligonucleotide microarray means that there are also windows of a much smaller size (down to ~0.5Kb, green windows) and windows of a larger size (up to ~4Kb, red windows). If we instead pass a sliding window of a fixed genomic size across the region, we avoid this problem (lower panel).

based on windows of a given genomic size rather than a given number of probes (*Figure 2.5B*). This was done using a script in R and by specifying the window size to be examined and the distance the window was to increment. The increment size was determined by the computing time required to complete the analysis, smaller window sizes obviously took more time.

2.12.4 Analysis of Correlation at Intervals of 1-100Kb

I wanted to examine the correlation between running means of chromatin structure datasets. However, rather than assume that a correlation may occur at any particular size of interval, I elected to scan many sizes of genomic interval. Therefore, a sliding window 1Kb in size was passed across the region and the mean value for each of the two datasets was calculated in each window (for example from Open/Input values resulting from sucrose gradient analysis and H3K9Ac/Input values resulting from ChIP). After the mean values were calculated for each window, the correlation between the two profiles was calculated and stored along with the window size. Then the positions of running mean windows of 2Kb in size were calculated and the process repeated again for 3Kb, 4Kb, etc all the way up to a 100Kb window size. The windows were calculated with an increment of a 10^{th} of the window size between each position due to the limits of computer processing power.

I then plotted the correlation against scale (*ie* sliding window size) which allowed me to identify interval sizes of interest between each pair of datasets where the correlation was greatest, the peak window size. To assign p-values to these peak values, I generated 1000 random samples of each dataset. I then generated running means of these paired random samples using a window size equal to that of the peak window size in the observed data and calculated the correlation of the running mean each time. The p-value of the peak of interest was therefore the percentage of random sample pairs whose correlation was greater than or equal to the correlation observed at the peak (for a positive correlation, in the case of a negative correlation

peak the p-value was the percentage of random sample pairs which had a correlation, less than or equal to the peak of interest).

2.12.5 Power Spectrum Analysis of Sucrose Gradient Data

To perform a power spectrum analysis, data has to be sampled over regularly spaced intervals. In physics this is usually over time but for the purpose of my analysis this would be over genomic distances. As explained above, the data on a tiling microarray is not sampled over regular intervals. Therefore the first step in performing the analysis was to model the data onto a regular interval of 1Kb. To do this I used the modelling functions that are available in R. I constructed a model of the $\log_2$ Open/Input values based on genomic position using a loess algorithm. Using this model I predicted, $\log_2$ Open/Input values sampled every 1Kb.

Loess modelling has the advantage that it makes no assumptions about any of the properties of the data. It instead calculates a modelled value for each data point based on its observed value and the value of datapoints close to it (in this case the value of probes that are located close to the probe of interest in the genome). Datapoints closer to the point of interest lend more weight to the calculated value than those lying further away. The main parameter that can be adjusted in the loess function is the span, α, which determines what percentage of datapoints are used in the calculation of each modelled datapoint. For most applications, this is generally set at around 0.5 to 0.75 (50-75%), resulting in a smoothed representation of the data. However, in this case I wished to closely model the original data, so I empirically determined a value which achieved that, 0.0025.

Having modelled the data, I next performed the power spectrum analysis. Fourier transformation of the data was performed using the fast Fourier transformation algorithm in R. This resulted in a series of frequencies with a corresponding amplitude value. The frequency is obviously the reciprocal of the period and has the units Kb^{-1} in this case. However, the data outputted by the R program in this analysis was in the form of complex numbers. To enable these

values to be plotted on a graph, I calculated their modulus (were a complex number is composed of a real part, *r*, and an imaginary part, *bi*, the modulus, $M = r^2 + b^2$, *i* is the imaginary unit with a value, $i^2 = -1$).

The observed amplitude peaks for each frequency correspond to the pattern of data but, not all of these will be biologically relevant. I assessed their relevancy by assigning p-values to each amplitude peak using a method adapted from C. Semple (originally based on (Higasa and Hayashi, 2006)). This method differs from more simple analyses by the use of randomised datasets to assess the significance of any observed frequency peaks. He demonstrated that this approach could be used to demonstrate the high significance of a three base pair periodicity in the genome of the bacteriophage Φ-X174 due to coding sequence (using the data from a previous study (Berger et al., 2003), C. Semple personal communication).The high density tiled region across the EDC on my microarrays consists of 3116 probes. I generated 1000 randomised data sets by replacing their $\log_2$ open/input values with a sample of 3116 values taken from the whole array dataset (~5000 values). I processed, these randomised datasets as for the observed dataset, ie by loess modelling them and then Fourier transforming them.

Therefore, for each frequency, I had a set of 1000 amplitude values derived by randomisation of the data, the control dataset for that frequency, and one the observed value. The biological relevance of the observed value could therefore easily be calculated by examining its position relative to the distribution of the control dataset. This was done as a one-sided statistical test, *ie* only control values greater than the dataset median were considered. The p-value of the observed amplitude was, therefore, the percentage of control values greater than or equal to that of the observed value. The p-value was calculated for each frequency, allowing significant periodicities in the data to be identified (using a cutoff of $p<0.01$).

2.13 Websites for Computational Biology

The bioinformatics and statistical resources referred to in this thesis are listed below, together with the relevant World Wide Web (www) link:

Ensembl	http://www.ensembl.org/
R-Project	http://www.r-project.org/
UCSC Browser	http://ucsc.genome.edu/
OMIM	http://ww.ncbi.nlm.nih.gov/omim
GEO	http://ww.ncbi.nlm.nih.gov/geo

Chapter 3: Characterisation of Gene Expression at the EDC in Model Cell Lines

In this project I wish to use transformed cell lines to investigate how chromatin regulates transcription at the human EDC. The first stage of such a study is to characterise cell lines with regard to their gene expression status at the locus to identify model cell lines in which the EDC is active and control cell lines where the EDC is silent. Comprehensive studies of gene expression in primary keratinocytes have been undertaken, for example see (Gazel et al., 2003;Mehul et al., 2004), but, there has been no survey of transformed keratinocytes. Therefore, I have undertaken my own analyses using RT PCR and expression microarrays to select a suitable model keratinocyte cell line. I also analysed EDC gene expression in a number of control cell lines using the same methods. Finally, I have differentiated a model keratinocyte cell line *in vitro* to determine what transcriptional changes were induced at the EDC.

3.1 RT PCR Analysis of Genes of the EDC

The gene clusters of the EDC are expressed in development in an order determined by the program of epidermal differentiation (Mischke, 1998), *Figure 1.5*, p40). To determine which developmental stage our model cell lines best represented, RT-PCR primers were designed for representative genes from within each of following EDC gene families; *S100A*, *fused*, *SPRR* and *LCE*. The genes selected were: *S100A6* and *S100A10*, which lie in the telomeric and centromeric *S100A* clusters respectively, the fused gene *Profilaggrin*, *SPRR1B* and *LCE3D* (*Figure 3.1A*). In addition, primers were designed for two potential house keeping genes that lie to either side of the EDC: *Sorting Nexin 27* (*SNX27*) and *Interleukin Enhancer-Binding Factor 2* (*ILF2*) (*Figure 3.1A*). *Beta-Actin* was also used as a loading control (primers kindly donated

by Jeremy Sanford). The primer pairs were designed to yield products of ~100 to 200 base pairs and to span an intron, thereby allowing the presence of contaminating genomic DNA in RT PCR reactions to be identified.

RT PCR was performed on RNA from 3 different keratinocyte cell lines, SVK14s, HaCaTs and NEB1s (*Figure 3.1B*). HaCaT cells are a spontaneously immortalized keratinocyte cell line (Boukamp et al., 1988). SVK14 and NEB1 cells are human keratinocytes that have been transformed by SV40 infection and the early region of human papillomavirus 16, respectively (Morley et al., 2003;Taylor-Papadimitriou et al., 1982). Each of the three keratinocyte cell lines expressed *S100A6* and *S100A10* at roughly equivalent levels (*Figure 3.1B*). *Profilaggrin* mRNA was also detected in all three of the cell lines, but at a lower level in SVK14 cells when compared to HaCaTs and NEB1s. Finally, no expression of *SPRR1B* or, *LCE3D* mRNAs was detected in SVK14 cells. *LCE3D* expression was confined to NEB1 cells, whereas *SPRR1B* was expressed by both HaCaT cells and NEB1 cells.

Assuming that the expression status of the single genes assayed here reflects the expression status of their parent clusters, we can draw conclusions about the relative stages of differentiation of the three cell lines. It is known that the *S100A* genes and *Profilaggrin* are expressed earlier in differentiation than the *SPRR* genes (Mischke, 1998), and the *LCE* genes are expressed only during the very final stages of differentiation (Marshall et al., 2001). The SVK14 cell line only expresses the earlier gene clusters, the *S100A* and *fused* genes, and therefore, represents an early differentiation state. On the other hand, all the EDC genes assayed are expressed in NEB1 cells and these cells may represent a late stage of keratinocyte differentiation. HaCaT cells express all the genes except *LCE3D* and are representative of a stage of differentiation intermediate between the SVK14 and the NEB1 cells. This analysis suggests that SVK14 cells would not make a good model for this study, because out of the three cell lines analysed they express the fewest genes in the EDC. In addition, these cells may have lost some of their epidermal identity and ability to differentiate into squames (Taylor-Papadimitriou et al., 1982). Both HaCaT and NEB1 cells express a number of the assayed genes, suggesting that they would make

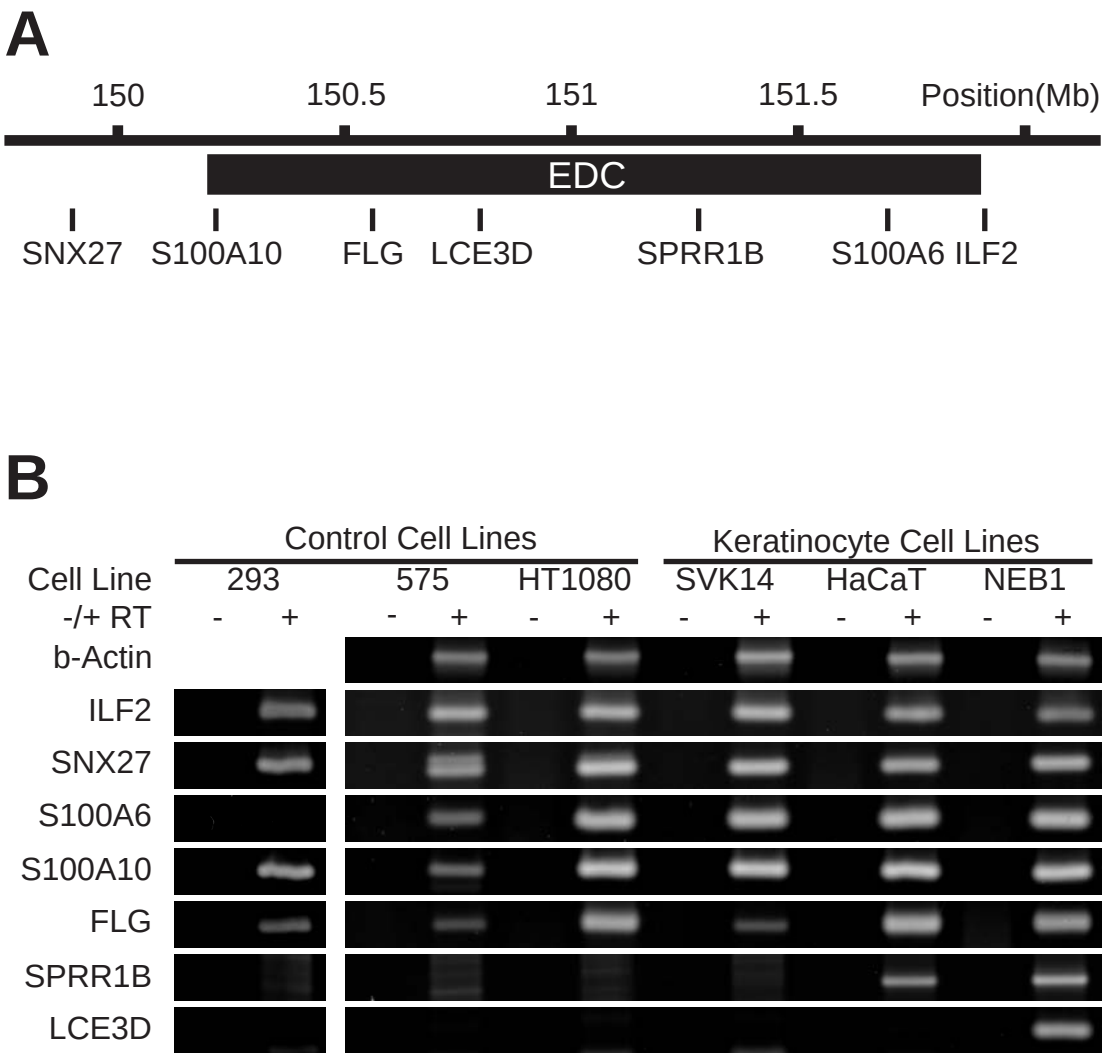


Figure 3.1 RT-PCR Analysis of Model Cell Lines

A. Schematic of the Human EDC. The position of each gene assessed by RT PCR is indicated by its midpoint on the NCBI Build 36 assembly on the UCSC Genome Browser (March 2006). The only gene not shown is b-Actin which lies on human chromosome 7.

B. RT PCR Analysis of Genes in Control and Keratinocyte Cell Lines. -RT is a control reaction carried without reverse transcriptase. No RT-PCR for b-Actin was carried out on cDNA from the 293 cell line.

good model cell lines with which to pursue further study. HaCaT cells are a well known and characterised model cell line that, importantly, retains full epidermal differentiation capacity (Boukamp et al., 1988). NEB1 cells on the other hand are relatively uncharacterised and it is not known how many epidermal characteristics they possess. Therefore, initially at least, it would seem prudent to use HaCaT cells as my keratinocyte model cell line.

RT PCR was also performed on a number of control cell lines (*Figure 3.1B*). Genome wide expression data from Nick Gilbert suggested that the EDC was inactive in lymphoblastoid cells (Gilbert et al., 2004) and therefore I selected 575 lymphoblastoid cells as one of my control cell lines. Published genome wide expression data suggests that the EDC is inactive in 293 cells (Staege et al., 2004), Gene Expression Omnibus (GEO) data set identifiers: GSM31805 and, GSM31806) and, they were also selected as a control cell line. Finally, I also assayed gene expression in HT1080 cells. They are a human fibrosarcoma cell line and it was thought that they were unlikely to express the tissue specific genes present in the EDC (Rasheed et al., 1974).

RT PCR analysis of the three control cell lines revealed that the two putative housekeeping genes, *ILF2* and *SNX27*, are expressed in every cell line assayed (*Figure 3.1B*). It was expected that the control cell lines would not express any of the genes in the EDC. However, bands were observed for the *S100A* genes and *Profilaggrin* in all three of the cell lines. *S100A6* was not expressed in 293 cells but *S100A10* was expressed to a level equivalent to that found in the keratinocytes. *Profilaggrin* was also weakly expressed in 293 cells. Weak expression of all three genes was found in 575 cells. On the other hand HT1080 cells expressed these three genes robustly and the level of *profilaggrin* expression was higher than the SVK14 keratinocyte cell line. However, the later EDC genes, *SPRR1B* and *LCE3D*, were not expressed in the control cell lines.

The fact that HT1080 cells express more of the assayed EDC genes than one of the keratinocyte cell lines suggests that in these cells there may a high amount of transcriptional activity in the EDC. Therefore, I conclude that they would be unsuitable for use as a control cell line to examine the chromatin structure of an inactive EDC. Both the 293 cells and 575 cells express fewer genes in the EDC and

would therefore be more suitable as controls. It would be preferable to find a cell line in which the EDC was completely inactive. However, it would seem unlikely that such a cell line exists. Firstly, I have briefly examined the transcription status of the EDC in a number of different cell lines using data from the Genomics Institute of the Novartis Research Foundation's survey of expression in a number of human cell lines (GNF gene atlas, GEO ascension GDS181, (Su et al., 2002). None of the cell lines examined appeared to have an EDC that was more repressed than that of 293 cells (data not shown). Secondly, I examined the expression status of the EDC in HEPG2 hepatoma cells and RITVA primary fibroblasts by RT PCR and found that they expressed a similar number of genes to the lines already tested (data not shown).

The main transcriptional activity in the EDC I found in my control cell lines appears, from this analysis, to be from the *S100A* genes (*Figure 3.1B*). Expression of these genes in cells other than keratinocytes is not unexpected because it is known that the S100A proteins are involved in a number of cellular processes including contraction, motility, cell growth, differentiation, cell cycle progression, transcription and secretion (Marenholz et al., 2004). This contrasts with the majority of other genes in the EDC, which are only known to be involved in epidermal differentiation. Data from the GNF gene atlas shows that the two *S100A* genes assayed by RT PCR here, *S100A6* and *S100A10*, are expressed in many tissues, eg. blood, adipocytes, cardiac myocytes and smooth muscle (Su et al., 2002), probe numbers 217728_at and, 200872_at respectively). In addition to their widespread expression in normal tissues, several of the *S100A* genes have roles in cancer. Like the cell lines used here, cancer cells are transformed, and it is possible that the simple act of culturing transformed cells results in misexpression of some of the *S100A* genes.

Unlike the S100A proteins, filaggrin is a highly specialised protein that is only known to function in the epidermis. Furthermore, out of the panel of tissues and cell lines examined by the GNF gene atlas, *profilaggrin* is only expressed in normal human skin (Su et al., 2002) probe number, 215704_at). It is difficult to imagine that it would be required in neuronal cells, lymphoblastoid cells or fibrosarcoma cells. Therefore, the expression I see in these cells by RT PCR is likely to be ectopic (*Figure 3.1B*). The fact that *profilaggrin* is expressed ectopically suggests that perhaps the regulation mechanisms acting on this gene are slightly less robust than

those acting on some of the other genes with the EDC. The only evidence thus far of specialised regulatory elements located within the EDC is from a study of the *SPRR* cluster (Martin et al., 2004). In agreement with this study, I do not see expression of the SPRR genes in my control cell lines (*Figure 3.1B*).

3.2 Whole Genome Expression Microarray Analysis of the EDC

Although RT PCR analysis has provided data on the expression status of 5 of the genes in the EDC, this is only ~1/10th of the whole locus. To be taken to gain a more complete picture of the level of transcriptional activity within the EDC, I elected to study global gene expression in the model cell lines using whole-genome oligonucleotide expression microarrays (kindly provided by B. Ylstra, see van den Ijssel et al., 2005).

There are two main designs for a microarray experiment utilising two-colour microarrays of this type. The first approach is to hybridise different experimental samples directly against each other. Alternatively, each experimental sample is hybridised against a reference sample (Churchill, 2002). I elected to use the second design for two reasons. Firstly, using a reference sample for all comparisons makes it easier to expand the study at a later date, by including another cell line for example (Churchill, 2002). Secondly, the control cell lines I use are widely used within our laboratory and data detailing their expression status, that is independent of a particular experiment, represents a useful resource for the lab. I chose to use the Stratagene Universal Human Reference RNA, a mixture of total RNA from 10 different human cell lines, as my reference sample.

I isolated 4 independent RNA samples from the three cell lines that were selected using my RT PCR data (293, 575 and, HaCaT cells). The samples were then reverse transcribed and labelled alongside an equal quantity of the reference RNA. Although 4 independent samples were hybridised from each cell line in the experiment, two samples from each cell line were often labelled and hybridised on the same day making the actual number of true biological replicates for each cell line

two (Churchill, 2002). To control for the differential incorporation of the dyes used to label the hybridised samples (Cyanine 3 and 5, Cy3 and 5), replicate hybridisations were performed in alternating dye orientations. That is to say, the RNA samples from each cell line were labelled an equal number of times in Cy3 and, Cy5, this is known as a dye swap (Churchill, 2002). Two control hybridisations, where the reference RNA was labelled in both colours and hybridised against itself, were also performed (a self hybridisation). Control hybridisations could easily be distinguished from experimental hybridisations by the even yellow colour of each spot (*Figure 3.2A*).

Data was extracted from scanned images of the hybridised microarrays using the Bluefuse program. In order to compare replicate hybridisations to each other and to compare data between different cell lines, the data had to be normalised. I elected to use a block lowess normalisation strategy for each hybridisation because it is capable of correcting both for intensity dependent artefacts in $\log_2$ ratios (deviation from a mean ratio $\log_2$ ratio of 0 for low and high intensity spots) and for systemic spatial variation in the array (caused by, for example, slight imperfections in the slide surface or local variations in hybridisation conditions on the slide, (see *Section 2.9.6*, p92 and Quackenbush, 2002). The use of a local normalisation algorithm requires that there are a large number of spots within each printed block of the array and that probes are arranged at random on the slide (they are not printed in related groups), conditions that the microarrays used in this study satisfy (van den Ijssel et al., 2005). After normalisation, a replicate filtering strategy, based on that described by Quackenbush, was used to remove probes from the analysis if their ratios were inconsistent across the two dye swap orientations (see *Section 2.12.1*, p100 and Quackenbush, 2002). Then, a mean $\log_2$ cell/reference value was then calculated for each probe on the arrays.

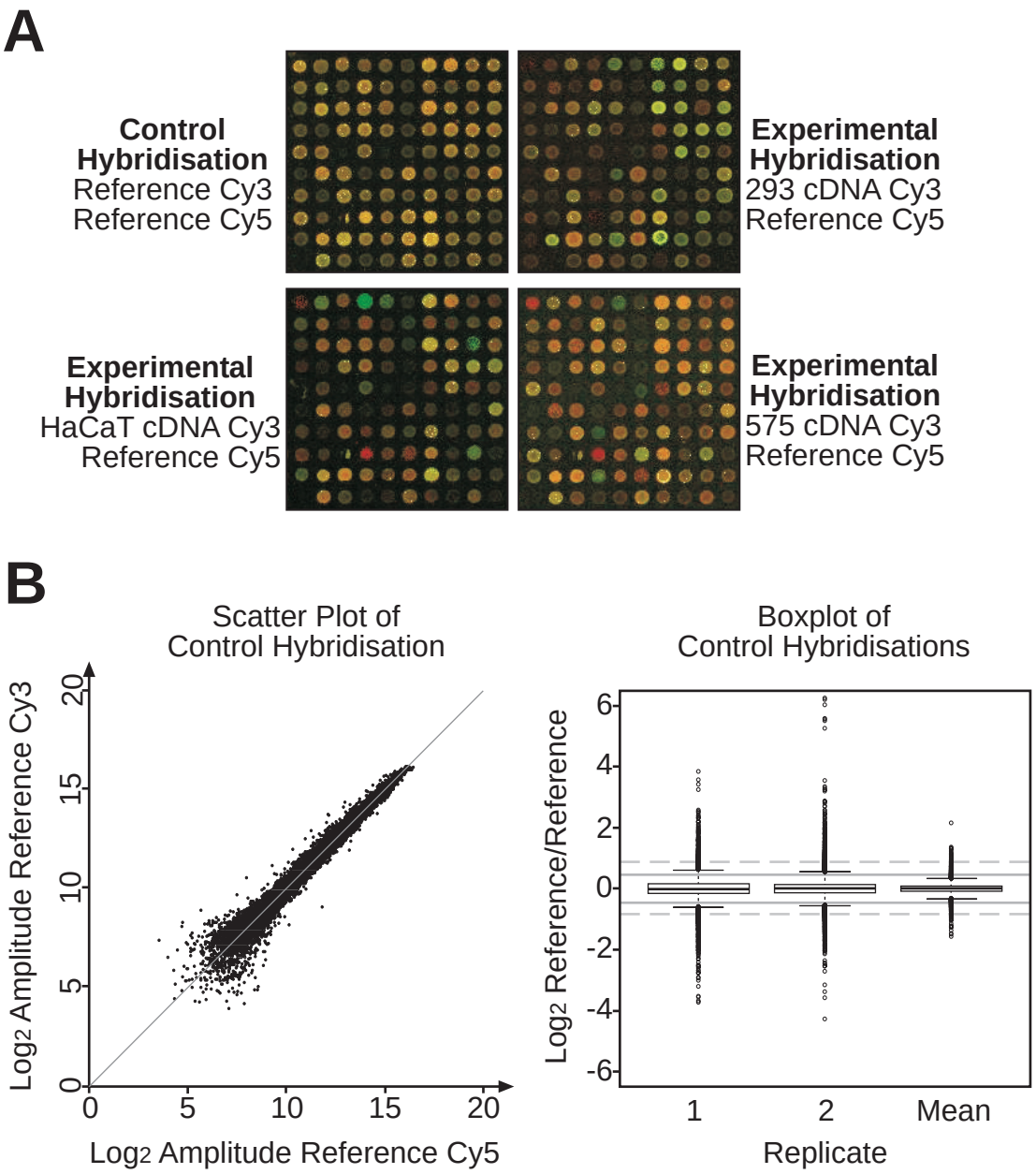


Figure 3.2 Microarray Analysis of RNAs from Cell Lines

A. Scans of microarrays hybridised with reference RNA and RNAs from model cell lines. Cy3 is shown as Green and Cy5 as Red. The same area of the array is shown in each picture.

B. Graphs of data obtained from control hybridisations. In the scatter plot, normalised $\log_2$ intensities in both channels were plotted for every spot from one hybridisation. The grey line indicates a Cy3/Cy5 Ratio of 1. In the boxplot, the distribution of normalised $\log_2$ ratios for each replicate hybridisation was plotted alongside the distribution of the mean $\log_2$ values from both replicates. The dashed grey lines indicate the 99% confidence level for one replicate and the solid lines the confidence level from the mean of the two hybridisations.

3.3 Estimation of Microarray Sensitivity for Analysis of Single Spots

One method to estimate the error inherent in any microarray experiment is to use control self hybridisations (see *Section 2.12.2*, p103 and Neal and Westwood, 2006). As expected, the ratios observed in each of my control hybridisations were scattered around 1 (or a log2 ratio of 0, indicated by the gray line in *Figure 3.2B*). However, the degree of scatter away from a $\log_2$ ratio of 0 in the control hybridisations was clearly lower than that observed in an experimental hybridisation (data not shown). A false positive rate (the probability that the $\log_2$ value for a particular oligonucleotide in an experimental hybridisation is equal to 0) can be estimated using the distribution of $\log_2$ values in a control hybridisation. By analysing the two control hybridisations individually I found that on average only 1% of the $\log_2$ values are greater than 0.886 and similarly only 1% are less than -0.842. Therefore, in a single experimental hybridisation I can have 99% confidence that an oligonucleotide with a $\log_2$ value outside of this range represents a true deviation from zero.

By considering multiple hybridisations together we can increase the sensitivity of the experiment. A boxplot showed that the mean $\log_2$ values from two control hybridisations deviated from zero less than their counterparts from either of the individual hybridisations (*Figure 3.2B*). The 99% confidence intervals for an up- or down-regulation were then calculated from the mean data in the same manner as before and found to be -0.468 and 0.466 respectively. Actual confidence levels in the experimental results will be higher than this value because 4 replicate hybridisations were performed for each cell line and the results should therefore be more accurate. Rough fold change estimates from my data which is in duplicate are 99.5% for a 1.5 fold and 99% for a 1.4 fold change. This compares favourably with previous studies which gave a 99.5% confidence estimate for a 1.5 fold change from quadruplicate data and, a 99% estimate for a 1.3 fold change from triplicate data (Johnston et al., 2004;Neal et al., 2003).

3.4 The EDC is Upregulated in HaCaT Cells

The array expression data can be used to test whether the EDC is activated in the HaCaT cells, as my RT PCR data suggests. Mean $\log_2$ cell RNA/ reference RNA values were ordered based on their annotated position (as supplied by B. Ylstra, van den Ijssel et al., 2005). 2414 probes had no annotated position and were removed from the analysis. Oligonucleotides corresponding to genes positioned in the EDC could then be identified (44 in total). Histograms were plotted to compare the distribution of mean $\log_2$ cell RNA/reference RNA values for the whole genome to that for oligonucleotides corresponding to the EDC (*Figure 3.3A*). Firstly, no differences were seen in the distribution of $\log_2$ values for the EDC and the whole genome in the data from the control hybridisations. This indicates that there are no specific errors in the measurement of these probes against the whole genome. The distribution of $\log_2$ values for the EDC also appeared to be the same as that for the whole genome in 293 and 575 cells. We can not assume that the distribution of $\log_2$ values from the EDC is normal so I used the distribution-independent Kolmogorov Smirnov test (KS-test) to statistically test these observations. No significant differences were found between the distributions of the whole genome and the EDC for the control hybridisations, 293 cells and, 575 cells (*Figure 3.3A*).

On the other hand, the distribution of $\log_2$ values for the EDC from HaCaT cells is shifted significantly rightwards relative to that for the whole genome (*Figure 3.3A*, the medians were 0.440 and -0.010 respectively) This confirms that, as expected, the EDC is upregulated in HaCaT cells compared to the reference RNA pool. KS-tests were also carried out to check that there was no significant upregulation of the entire of chromosome 1 in HaCaT cells. The distribution of log2 values for chromosome 1 was significantly different from that of the whole genome, but comparison of the distributions showed that their was no easily identifiable up or down regulation (*Figure 3.3B*). No significant differences were seen in the distributions of chromosome 1 and the whole genome in either 293 or 575 cells (data not shown).

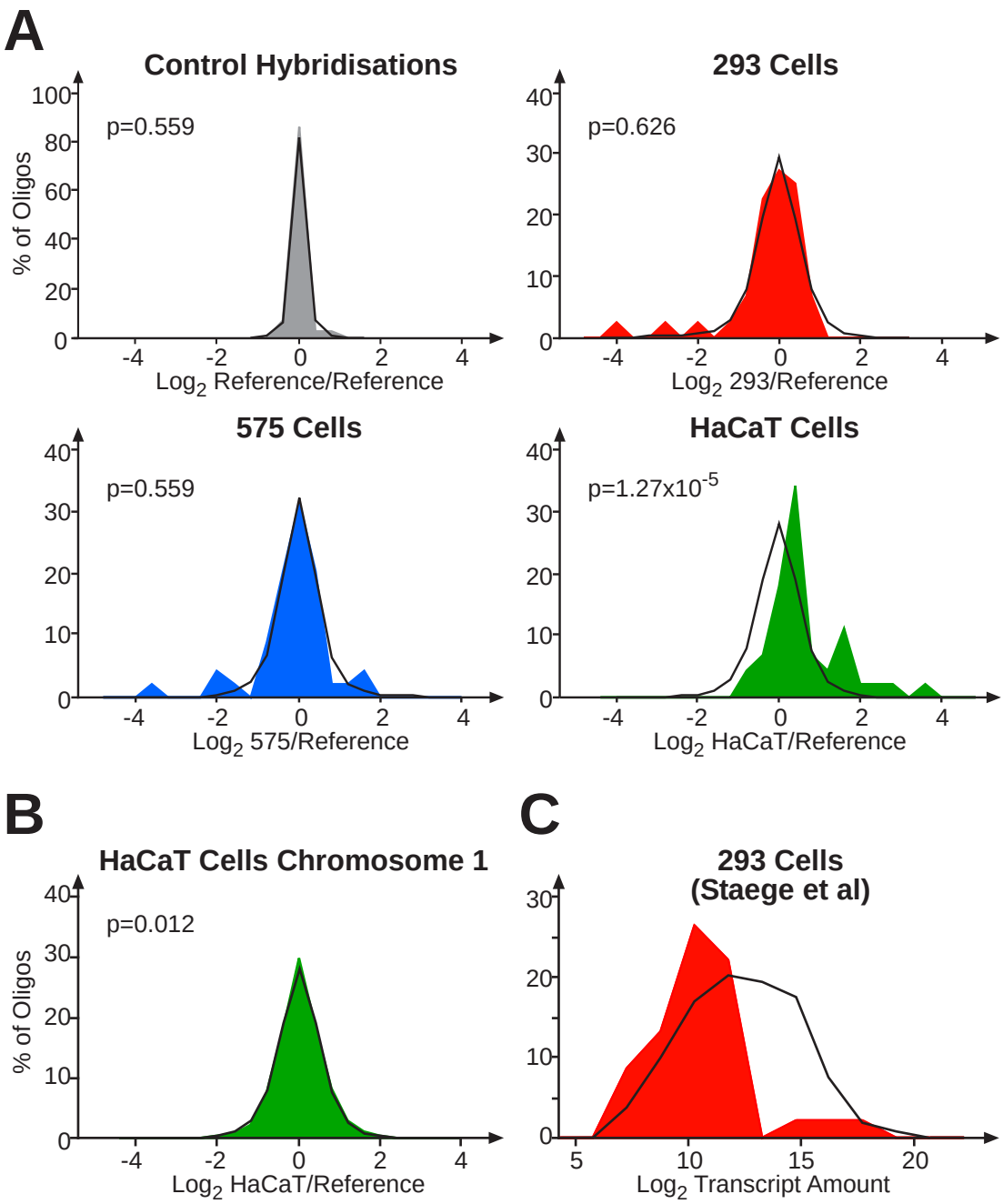


Figure 3.3 The EDC is Upregulated in HaCaT Cells

A. Histograms showing the distribution of array expression data. The distribution of $\log_2$ values of all oligos on the array is indicated by the black line and the distribution of $\log_2$ values for the EDC by the solid coloured histogram. The p-value of the KS-test between the two distributions is indicated in each case.

B. As in A, except the coloured histogram represents distribution of $\log_2$ values for chromosome 1 rather than the EDC. The p-value of the KS-test between the two distributions is indicated.

C. Histogram of the data from Staege *et al*., 2002 for 293 Cells. Absolute transcript levels were $\log_2$ transformed and a histogram plotted for all probes on the array (black line) and the oligos representing the EDC (red histogram).

I further tested the significance of the upregulation of the EDC seen in HaCaT cells using data from the control hybridisations. A running median was applied to the ordered control data to generate the median of every group of 44 probes across the whole array. No group of 44 adjacent probes had a $\log_2$ median $\geq$0.440 (the median $\log_2$ value of the EDC in HaCaT cells). As a further test, the control data was randomly permutated 1000 times and each time a running median was passed over each set of rearranged data. Even when this was done, no groups of 44 adjacent probes were generated that had a median $\geq$0.440. Therefore, the probability that the median $\log_2$ value observed for the EDC in HaCaT cells was generated by error and that the true value is 0 is vanishingly small. Lastly, the expression data from HaCaT cells was examined to determine if regions other than the EDC were upregulated. A 1.71Mb sliding window (the size of the EDC) was passed across the entire genome in 100Kb intervals. For each interval the median $\log_2$ value was calculated, generating a position based running median for the entire genome. 1236 (4.412%) of these intervals had a median $\log_2$ value $\geq$0.440. However, a large portion of these (338) represented the value of 1 or 2 oligonucleotides and therefore could be generated by one highly expressed gene. The EDC was among the list and is, therefore, among the most active genomic regions in HaCaT cells confirming my initial hypothesis.

As mentioned in *Section 3.1*, I chose to use 293 cells as a control cell line because my analysis of published data suggested that the EDC was inactive in these cells (Staege et al., 2004). I now compared that data to my own microarray data to provide external validation of my results. Staege and collegues used Affymetrix HG-U133A arrays to assay gene expression. I downloaded the data from the GEO service and identified probes belonging to the EDC using the Affymetrix HG-U133A data track on the University of California Santa Cruz Genome Browser (UCSC browser). A mean transcript level was calculated for every probe from the two replicates in the study and a histogram of $\log_2$ mean transcript amount for the EDC superimposed on a histogram for every probe on the array (*Figure 3.3C*). Examination of this histogram revealed that the distribution for the EDC was shifted to the left relative to the distribution of all probes. This suggests that the EDC is

repressed and is different from my own result where the two distributions are superimposed (*Figure 3.3A*).

This may be due to the differing experimental designs used. Affymetrix arrays measure absolute transcript levels of one sample rather than a ratio of two transcript levels. I used two colour microarrays which measure a ratio of transcript levels between two samples (in this case cell lines vs reference RNA). The reference RNA sample is composed of a mixture of total RNAs from 10 human cell lines but does not include RNA from a keratinocyte cell line (Universal Human Reference RNA Manual, Stratagene). Transcript levels of many of the genes in the EDC would be predicted to be low in the reference sample because expression of these genes is known to be highly restricted to keratinocytes. Therefore, if the majority of the EDC is inactive in 293 cells as suggested by the data of Staege *et al.* I would expect to see $\log_2$ ratios close to 0 (signifying no difference in abundance between 293 cells and the reference sample) for the majority of genes in the EDC. A prediction of this hypothesis is that if the control cell lines were directly compared to HaCaT cells (*ie* their RNAs were competitively hybridised), then the distribution of $\log_2$ ratios for the EDC would be skewed in the direction of the HaCaT cells. If I simulate a direct comparison using my data, this is indeed what I find (Median $\log_2$ HaCaT/293 = 0.527 and, $\log_2$ HaCaT/575 = 0.519 for the EDC). Therefore, I conclude that the majority of the EDC is repressed in my two control cell lines, 293 and 575.

3.5 Transcriptional Regulation of the EDC

Although the EDC was first defined as a gene cluster in 1996 there have, thus far, been no studies examining the expression of the whole region during epidermal differentiation (Mischke et al., 1996). The region has a complex organisation and is much larger and more gene-rich than many of the more studied gene clusters (for example, the *Hox* and *β-globin* loci). During epidermal differentiation, the *S100A* genes are the first to be expressed and the fused genes and *SPRR* genes are only activated as differentiation proceeds (Mischke, 1998). Finally, the *LCE* genes are activated in the very final stages of epidermal differentiation (Marshall et al., 2001). If this pattern of expression is superimposed onto a physical map of the EDC (*Figure*

3.4A), we can hypothesise that gene expression begins at the edges of the locus and then spread into the centre as differentiation proceeds.

In order to test this hypothesis, I mapped my genome wide expression data from HaCaT cells onto a physical map of the EDC (*Figure 3.4B*). This also enabled me to confirm that the genome wide expression data correlated with the RT PCR analysis of expression detailed in *Section 3.1* (see, *Table 3.1*). As noted above, there are difficulties interpreting how array expression data generated using a reference RNA relates to the absolute levels of a transcript. Both the housekeeping genes (*ILF2* and *SNX27*) have low positive $\log_2$ values suggesting that they are expressed, as does RT PCR data. However, if they are truly housekeeping genes then we would expect them to be expressed to a roughly equal level in all cells and the $\log_2$ ratio to be close to 0. This is because the RNAs should be present in a roughly equal level in all the cell lines that make up the reference RNA and the HaCaT cells. The fact that I see a positive $\log_2$ value suggests that either they are slightly upregulated in HaCaT cells or that they are not expressed or expressed to a lower level in some of the cell lines that make up the reference RNA. Support for this finding comes from examining data from the GNF gene atlas which shows that absolute transcript levels vary across tissues (Probe Identifiers: 200052_s_at for *ILF2*, 221006_s_at and, 221498_at for *SNX27*). *S100A6*, *S100A10* and, *profilaggrin* were all found to be expressed by RT PCR and they have high HaCaT/Reference ratios in agreement with this. *SPRR1B* was also determined to be active by RT PCR and has a much lower, but still positive, $\log_2$ ratio on the microarrays. Given that we expect this RNA to be absent from the reference pool, this analysis suggests that although the gene is expressed in HaCaT cells, it is expressed to a low level (the $\log_2$ ratio is greater than the 99% confidence intervals calculated in *Section 3.2*). Lastly, *LCE3D* has a $\log_2$ ratio that is very close to 0. We expect this RNA to be absent from the reference pool and, the HaCaT samples giving a $\log_2$ ratio of zero. Therefore, this result is in agreement with the lack of transcript as detected by RT PCR. These analyses suggest that the results of my RT PCR analysis and microarray expression analysis are correlated.

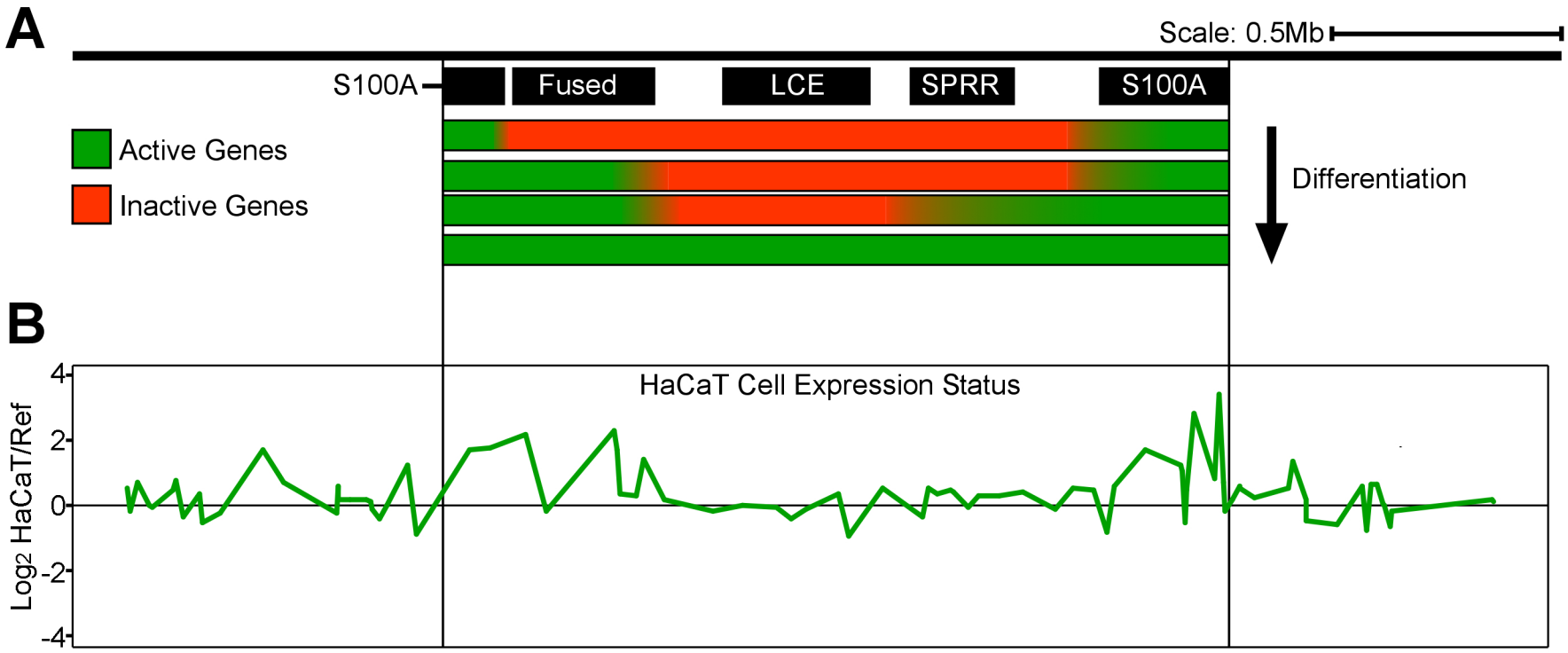


Figure 3.4 Mapping of Microarray Expression Data onto the EDC

A. Schematic of the proposed pattern of expression of the EDC. The region of the EDC is depicted along with the positions of its gene clusters. Below this a schematic shows the proposed pattern of gene activation during keratinocyte differentiation. Expression begins in the S100A genes and spreads towards the centre as differentiation proceeds.

B. Microarray Expression Data from HaCaT Cells overlayed onto a map of the EDC. The Log_2 HaCaT/Reference RNA values were plotted against the genomic position of each oligonucleotide as annotated in van de Ijssel *et al* 2005. The same annotation was used to determine the boundaries of the gene clusters.

The mapping of array data onto the EDC suggests that the hypothesis of the pattern of gene activation detailed above may be true. High $\log_2$ ratios, indicative of a higher level of expression, are restricted to the edges of the EDC (*Figure 3.4B*). This pattern resembles that of the hypothetical pattern for one of the earlier stages of epidermal differentiation (*Figure 3.4A*). However, not all of the members of a particular gene cluster are active at the same time as there are dips in the level of gene expression, even within highly active clusters like the *S100A*s. The map also shows that, in general the *SPRR* cluster is not particularly active in HaCaT cells. Therefore, analysis of the array data suggests that the HaCaT cell line represents an earlier stage of differentiation than I hypothesised from the RT PCR data. However, we should also bear in mind that these expression levels are relative to the reference sample and conformation of this pattern will require further quantitative studies.

Table 3.1

Comparison of RT PCR and Array Expression Data

Mean $\log_2$ HaCaT/Reference values were taken using the annotation of van den Ijssel *et al*.(2005) and compared to the RT PCR Data.

Gene Name	RT PCR Status	Mean Log2 Ratio
SNX27	ON	0.589
ILF2	ON	0.539
S100A6	ON	1.271
S100A10	ON	1.695
Profilaggrin	ON	1.742
SPRR1B	ON	0.487
LCE3D	OFF	0.025

3.6 Upon Differentiation, HaCaT Cells Upregulate the SPRR and LCE Genes

Keratinocytes, including HaCaTs, can be differentiated *in vitro* to recapitulate many of the features of the *in vivo* epidermal differentiation program (Boukamp et al., 1988;Paramio et al., 1998). Therefore, I differentiated HaCaT cells by using a published method (see *Section 2.6.4*, p73 and (Paramio et al., 1998). Briefly, high density cultures were kept in serum free media over the period of 20 days and total protein and RNA extracts taken at regular intervals. During this time the morphology of the cells changed dramatically (*Figure 3.5A*) and large quantities of specific proteins accumulated in the cultures (*Figure 3.5B*).

In order to characterise the specific changes at the transcriptional level occurring in the cells, I first assayed by western blot the three markers of epidermal differentiation described by Paramio *et al*., (Cytokeratin 10, Cytokeratin 14 and, Involucrin, *Figure 3.5C*). As reported, Cytokeratin 10 accumulated from day 4 of the differentiation onwards, but contrary to their report Cytokeratin 14 was not detected at day 0 and accumulated in the cells from around day 8. Unlike the Cytokeratins, Involucrin is encoded by a gene in the EDC and it was present at fairly constant levels in the cells throughout the 20 days of differentiation. This is a different expression pattern to that reported by Paramio and collegues who saw the protein appear in the cells from around day 12. To confirm that Involucrin is present in HaCaT cells before differentiation, separate protein extracts were made from cells obtained from two separate sources (B. Lane, as used so far in this study, and N.E. Fusenig). A specific band was observed in both in the extracts made from both clones of the cell line but not in extracts made from 293 cells (data not shown).

I also analysed the expression of these markers at the transcriptional level. RT PCR for *Cytokeratins 10* and *14* showed that there was no change in their mRNA levels during differentiation (*Figure 3.5D*). As expected from the western blots, the level of *involucrin* mRNA also did not change greatly across the time period. The presence of these mRNAs at day 0 of differentiation was confirmed by analysis of the array expression data ($\log_2$ HaCaT/Reference values: 2.883 for *involucrin*, 0.237

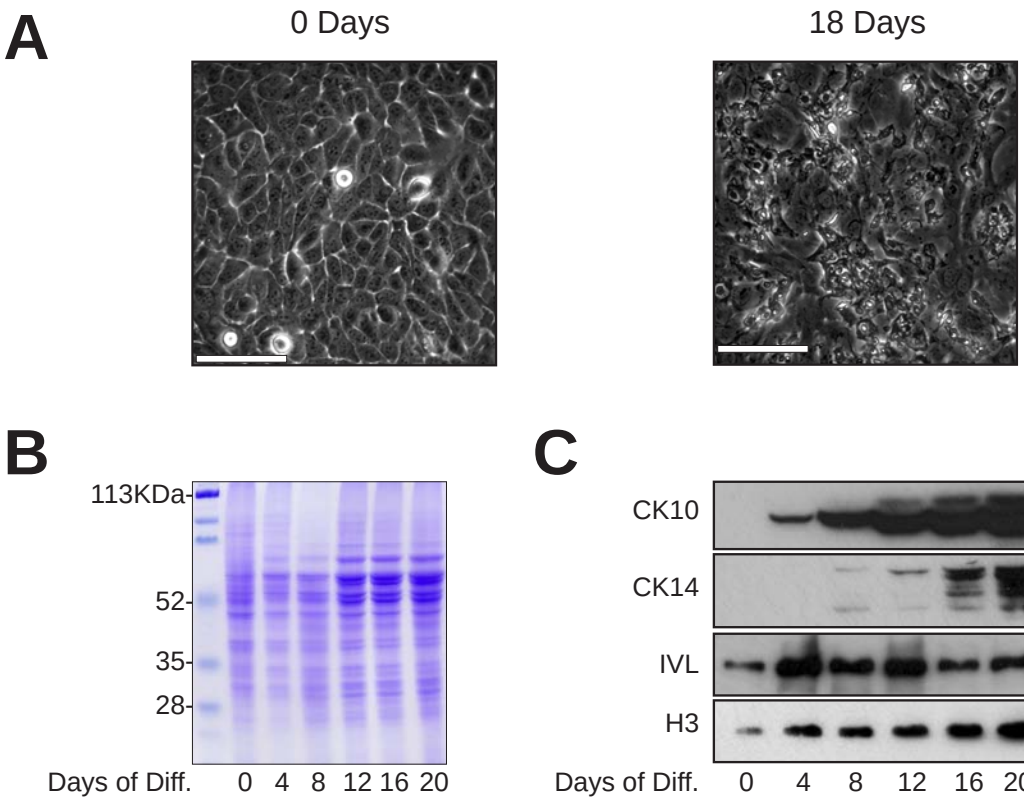


Figure 3.5 Transcriptional Regulation during HaCaT Differentiation

A. Phase contrast Microscope Pictures of HaCaT Cells before differentiation (Day 0) and after 18 days of differentiation in serum free media. The Scale bars represent 100 microns.

B. Coomassie stained SDS PAGE gel of whole cell extracts during the course of 20 days of differentiation.

C. Western blots of HaCaT whole cell extracts to detect Cytokeratin 10 (CK10), Cytokeratin 14 (CK14), Involucrin (IVL) and Histone H3 (H3).

D. RT PCR Analysis of gene expression during HaCaT differentiation. -RT is a control reaction carried out in the abscence of Reverse Transcriptase.

for *Cytokeratin 14* and, for 0.390 for *Cytokeratin 10*). The fact that RNA levels of the *Cytokeratins* do not appear to change when the levels of their protein products do suggests that the regulation of these genes during epidermal differentiation may be mainly at the post-transcriptional level. I also assessed the effects of differentiation on some of the other genes of the EDC. RT PCR for *SPRR1B* showed that, although it is present from day 0 of differentiation, its RNA accumulates in the cells as differentiation proceeds (*Figure 3.5D*). *LCE3D* is not expressed in undifferentiated HaCaT cells but was faintly detected after 4 days differentiation and more robustly after 8 days. Therefore, I can conclude that, as expected, these genes become activated as differentiation progresses.

If I use this RT PCR data as a crude assay for the expression status of the individual gene clusters in the EDC, the fact that *SPRR1B* and *LCE3D* are sequentially activated as differentiation proceeds provides evidence to support the model presented in *Figure 3.3A*. It would suggest that from around 4 days of differentiation the *SPRR* cluster becomes activated in HaCaT cells and, at around 8 days of differentiation the *LCE* genes are strongly expressed. However, I have already seen that the expression status of one gene in a cluster does not necessarily reflect the status of the entire gene cluster and this result will have to be verified with further microarray studies. In addition, it remains to be determined whether or not the early genes of the EDC, for example the *S100A*s, are repressed during the later stages of epidermal differentiation. There have been studies of the localisation of the S100A proteins in normal epidermis which suggest that they are still present in differentiated keratinocytes, but nobody has yet examined these genes at the RNA level (Boni et al., 1997;Broome et al., 2003;Sakaguchi et al., 2003). Because the *in vitro* differentiation of HaCaT cells produces a mixed population of differentiated and undifferentiated cells, it does not make a good model with which to answer this question (Boukamp et al., 1988). However, my RT PCR data suggests that NEB1 cells represent a late stage of epidermal differentiation and they still express *S100A6* and *S100A10* (*Figure 3.1B*), therefore, they may remain active as epidermal differentiation progresses.

3.7 Discussion

In order to study the role of chromatin in the transcriptional regulation of the EDC, it is important to first determine what transcriptional activity occurs at the EDC during differentiation. I have chosen to use immortalised cell lines as a model of epidermal differentiation. Although there have been a number of studies examining the expression of individual members of the EDC in cultured cells, there has yet to a study of the entire locus (Boukamp et al., 1988;Lapi et al., 2006;Lesniak et al., 2007;Paramio et al., 1998). I analysed the expression status of the EDC in HaCaT cells by RT PCR and expression microarrays. I found that, as expected, the EDC was significantly upregulated in these cells. No upregulation of the locus was seen in two control cell lines, 293 cells and 575 cells.

By mapping the array expression data onto the EDC locus, I concluded that HaCaT cells represent one of the earlier stages of epidermal differentiation because the later gene clusters of the EDC, the *SPRR* and *LCE* loci, are not generally active. Then I differentiated the cells *in vitro* and, showed that individual members of these later gene clusters were induced, suggesting that the entire clusters may be activated as differentiation proceeded. These two experiments combined support a model of gene activation at the EDC whereby the outer *S100A* clusters are expressed first and as epidermal differentiation progresses, gene activation proceeds inwards (*Figure 3.4A*). To now understand how the EDC is regulated, I can now go on to characterise the chromatin structure of the EDC in HaCaT and control cells.

Chapter 4 Interphase Chromatin Organisation of the EDC

I have established that the EDC is transcriptionally active in HaCaT cells, so next I wanted to determine whether these cells display any large scale differences in interphase chromatin structure when compared to control cell lines. Large scale chromatin structure can be assayed using cytological techniques such as FISH, for example, by determining the nuclear positioning of a genomic locus relative to a particular nuclear compartment or another genomic locus. A second assay analyses the large scale compaction of chromatin in a particular genomic region by measuring the distance between pairs of probes in the nucleus. Previously, these techniques have been used to show that, in general, gene-rich areas of the genome are cytologically decondensed and more often positioned away from their parent chromosome territories when compared to gene-poor regions (Gilbert et al., 2004;Mahy et al., 2002a;Mahy et al., 2002b;Yokota et al., 1997). However, it has also been shown that many co-ordinately regulated genomic regions, including the EDC, are positioned inside their parent chromosome territory when inactive but move outside when active (Chambeyron and Bickmore, 2004;Volpi et al., 2000;Williams et al., 2002). In addition, the *HoxB* and *D* loci have been seen to cytologically decondense upon induction of differentiation in mouse embryonic stem cells and in embryos (Chambeyron and Bickmore, 2004;Chambeyron et al., 2005;Morey et al., 2007). In this chapter, I have assayed large scale chromatin structure at the EDC in HaCaT cells and compared this data to what we already know about gene expression in the EDC. However, first I have examined the karyotypes of the cell lines in order to make sure that the EDC is not disrupted by chromosomal rearrangements.

4.1 The EDC is not Affected by Chromosomal Rearrangements in HaCaT Cells

Cell lines in culture often become anueploid and acquire chromosomal abnormalities. The EDC is a relatively large locus and may be affected by such rearrangements, therefore, it is important to establish that this has not occurred in my model cell lines. I investigated the integrity of the EDC by performing FISH on metaphase preparations. Three BACs were used as probes: one lying within the *SPRR* cluster (Probe 1, *Figure 4.1*), one lying just outside of the EDC towards the telomere (Probe 2) and, another which lies approximately 2.5Mb away from the EDC in chromosome band 1q22 (Probe 3). I also used a chromosome paint specific for chromosome 1 in order to unambiguously identify both normal and abnormal copies of the EDC's parent chromosome.

The three probes all hybridised to the correct location on metaphase chromosomes from 575 cells (*Figure 4.2*) and these cells appeared to have a normal, diploid karyotype as previously described (Woodfine et al., 2004). This result confirms both that the probes are specific for the EDC and that the EDC has not been subject to rearrangement in 575 cells. On 293 spreads, the chromosome 1 paint hybridised to 3 different chromosomes. Two of these were normal copies of chromosome 1 and the other appeared to be a derivative in which an unbalanced translocation had occurred close to the telomere of chromosome arm 1q (*Figure 4.2*). In some metaphases, two of these derivatives were observed along with one normal chromosome 1 (Shelagh Boyle, personal communication). Consistent with my observations, it has previously been reported that 293 cells are hypotriploid with 2 normal copies of chromosome 1 and one derivative (Goetze et al., 2007). The translocation must lie in the order of 80 or 90Mb from the EDC and is unlikely to affect the locus. I observed that all three probes hybridised correctly, both to the normal chromosomes and the derivatives and concluded that, indeed, the EDC appeared unaffected by this or any other arrangements. It has also been found that the interphase chromatin structure of a region adjacent to the EDC in 293 cells is apparently unaffected by the rearrangement (Goetze et al., 2007). Therefore, despite

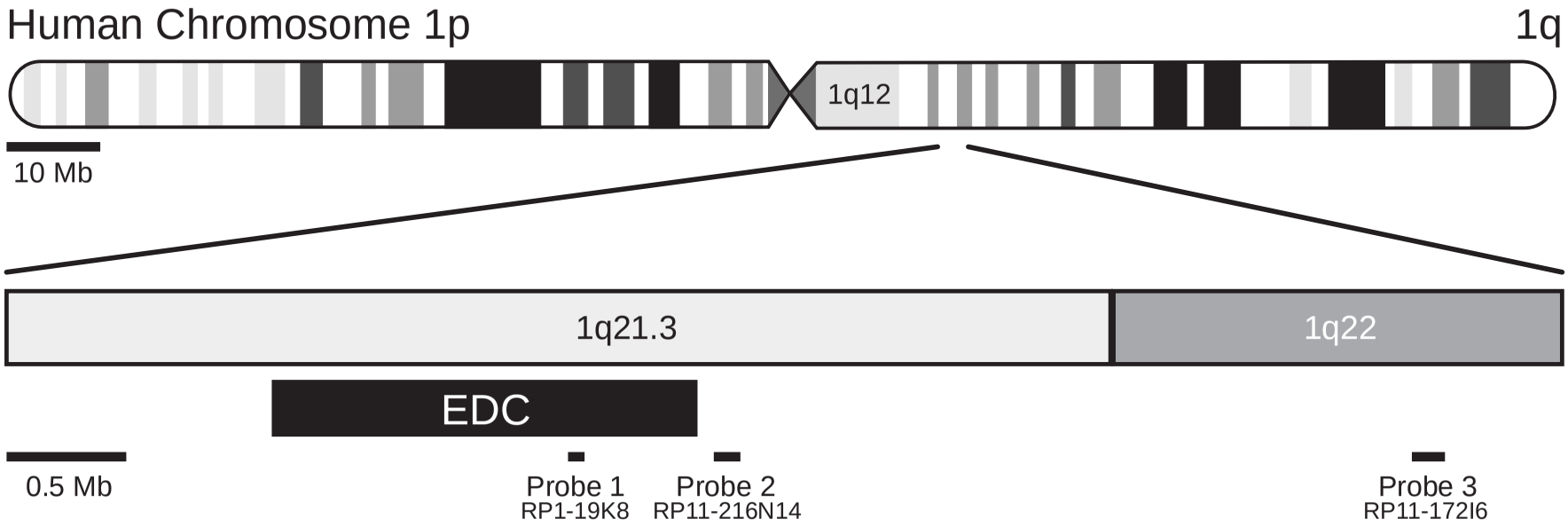


Figure 4.1 Human Chromosome 1 and the EDC showing location of FISH Probes

An ideogram of human chromosome 1 is shown with the p and q arms indicated along with the location of the EDC. The probes used in FISH are drawn roughly to scale and their universal names given below. For the genomic positions of each probe see *Section 2.7.9* p80.

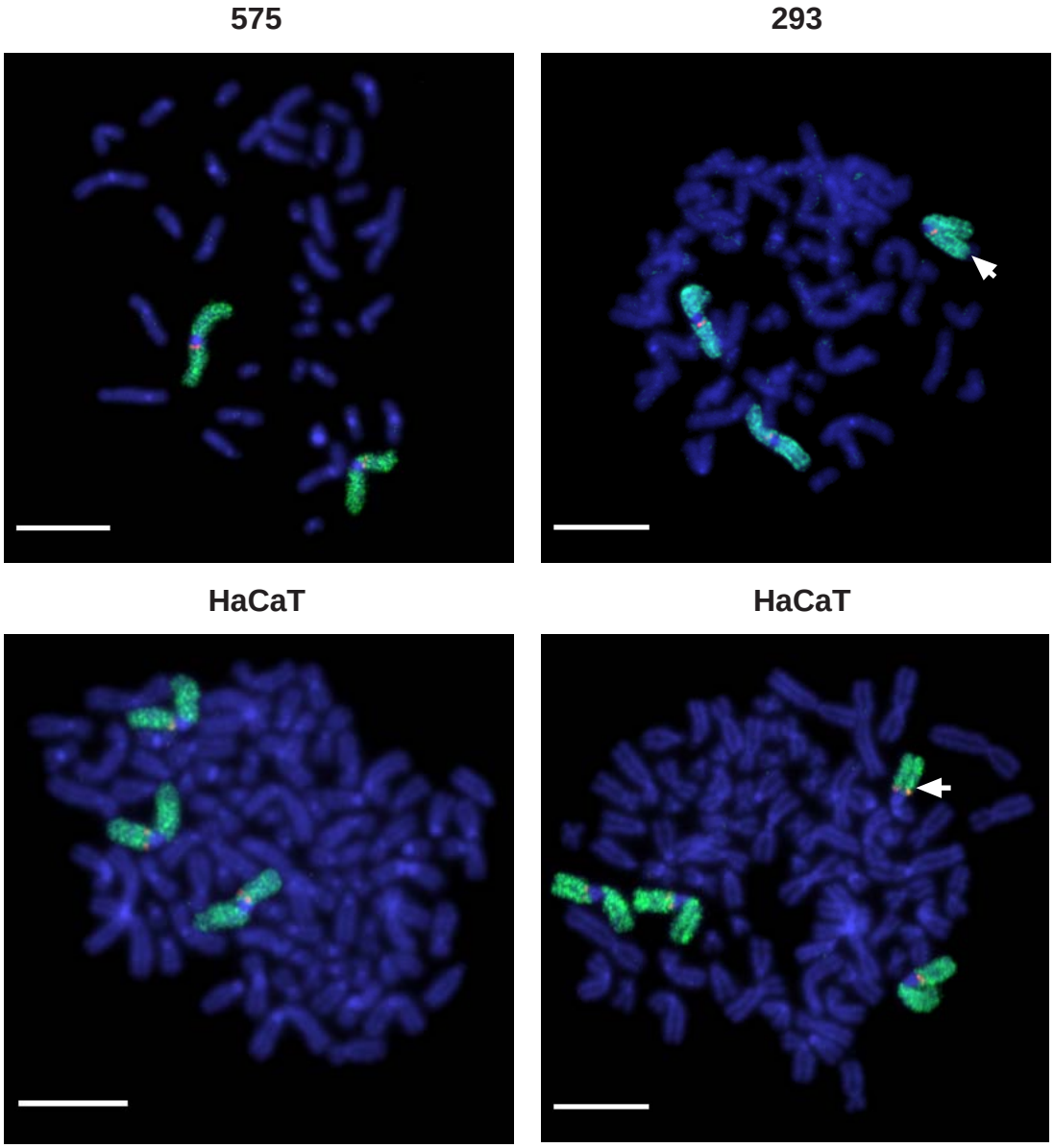


Figure 4.2 Karyotypes of Model Cell Lines

Images of metaphase spreads from 575, 293 and HaCaT cells hybridised with a chromosome paint for chromosome 1(Green) and Probe 1 (Red). The DNA has been counterstained with DAPI (Blue). On the 293 spread the arrow indicates the derative of chromosome 1 with an extension to chromosome 1q. On the second HaCat spread the arrow indicates an abnormal derivative of chromsome 1 where the p-arm appears to have been involved in an unbalanced translocation. The Scale bars indicate 10mm.

the presence of the abnormal derivative of chromosome 1 in this cell line, it is a good model with which to study the chromatin structure of the EDC.

I observed 4 normal copies of chromosome 1 in the majority of metaphases from HaCaT cells, consistent with the reported karyotype of the cell line at derivation (Boukamp et al., 1988). However, some metaphases had either three copies of chromosome 1 or one of the four copies was an abnormal derivative, with what appeared to be a truncation or unbalanced translocation involving most of chromosome arm 1p (*Figure 4.2*). Boukamp *et al.*, also reported that they observed chromosomal aberrations involving chromosome 1 in some metaphases. All the derivatives involved a truncation of the p arm of chromosome 1 and the three probes hybridised correctly to them and to the normal copies of chromosome 1, confirming that the EDC was unaffected by rearrangements in HaCaT cells. Similar analyses of two other model keratinocyte cell lines showed that both contained abnormal derivatives of chromosome 1. In SVK14 cells this involved duplication of the centromere and chromosome arm 1q and in NEB1 cells there was a duplication of the EDC and part of chromosome 1q with breakpoints lying between Probes 1 and 2 (Data not shown). These rearrangements both affect the portion of chromosome 1 containing the EDC and therefore confirm that these cell lines are unsuitable models for this study.

To study the chromosomal structure of the EDC in more detail, I performed array comparative genomic hybridisation (aCGH). This technique allows relative copy number differences between two genomes to be detected using genomic microarrays (Pinkel and Albertson, 2005). I am specifically interested in the genomic structure of the EDC, so I used custom-made oligonucleotide arrays that had been specially designed for me by Invitrogen in partnership with Illumina and printed at the University of Liverpool. They consisted of 5000 60-mer oligonucleotide probes, including 2145 located in the EDC with a resolution of roughly 1 probe per 800bp (for full details of the design of the array, see *Section 2.9.2* p90). Using these arrays, any disruption to the EDC can be mapped with a high degree of precision. It seemed likely that no disruptions had occurred within the region in 575 cells so, I elected to directly test whether there was any difference in the genomic structure of the region between HaCaT and 293 cells.

I isolated genomic DNAs from the two cell lines and fluorescently labelled them before hybridising them twice to the arrays in opposite dye orientations. From the metaphase FISH, I already know that in HaCaT cells there are 4 copies of chromosome 1 whereas in 293 cells there are only 3 copies and I therefore might expect an overall HaCaT/293 genomic ratio of around 1.333. However, the median HaCaT/293 genomic DNA ratio observed is close to 1 (data not shown). This is because the two fluorophores are separately scanned using a laser power and photomultiplier gain settings that give the greatest dynamic range of intensities in each channel (*ie* few saturated spots). Most spots on the array come from chromosome 1 and therefore the result of this process for these arrays is that on average the ratio of intensities between the two channels will be 1. To normalise the two replicates I used a lowess normalisation which has the ability to correct intensity dependent dye biases that occur during hybridisation (Quackenbush, 2002). I elected not to use the local normalisation strategy previously applied to analyse genome wide expression levels (see, *Section 3.2*, p115) because, although each printed block of the EDC arrays contains 462 spots, there are 4 replicates of each probe per block and therefore only 115 unique probes and I felt that this was not a sufficiently high number to reliably perform a localised, block lowess normalisation. After normalisation I derived average intensity values for each probe in the two hybridisations from its 4 printed replicates. A replicate filtering strategy was then used to remove any probes from the analysis whose ratios were inconsistent across the two replicate hybridisations (*ie* inconsistent across the dye swap) and finally a mean HaCaT/293 genomic DNA ratio was calculated for each probe from the two replicate hybridisations.

To test whether the relative copy number of the EDC was different between two cell lines, I compared the distribution of $\log_2$ HaCaT/293 ratios for oligonucleotides from the EDC to that of all other probes (*Figure 4.3A*). The distribution of the EDC probes is slightly shifted towards positive $\log_2$ values, indicating that the EDC may have a higher copy number in HaCaT cells relative to the rest of the genome. To test this further, I performed control self hybridisations to estimate the technical error of the assay. Two control hybridisations were scanned and normalised as described above. Because the DNA was self hybridised to arrays

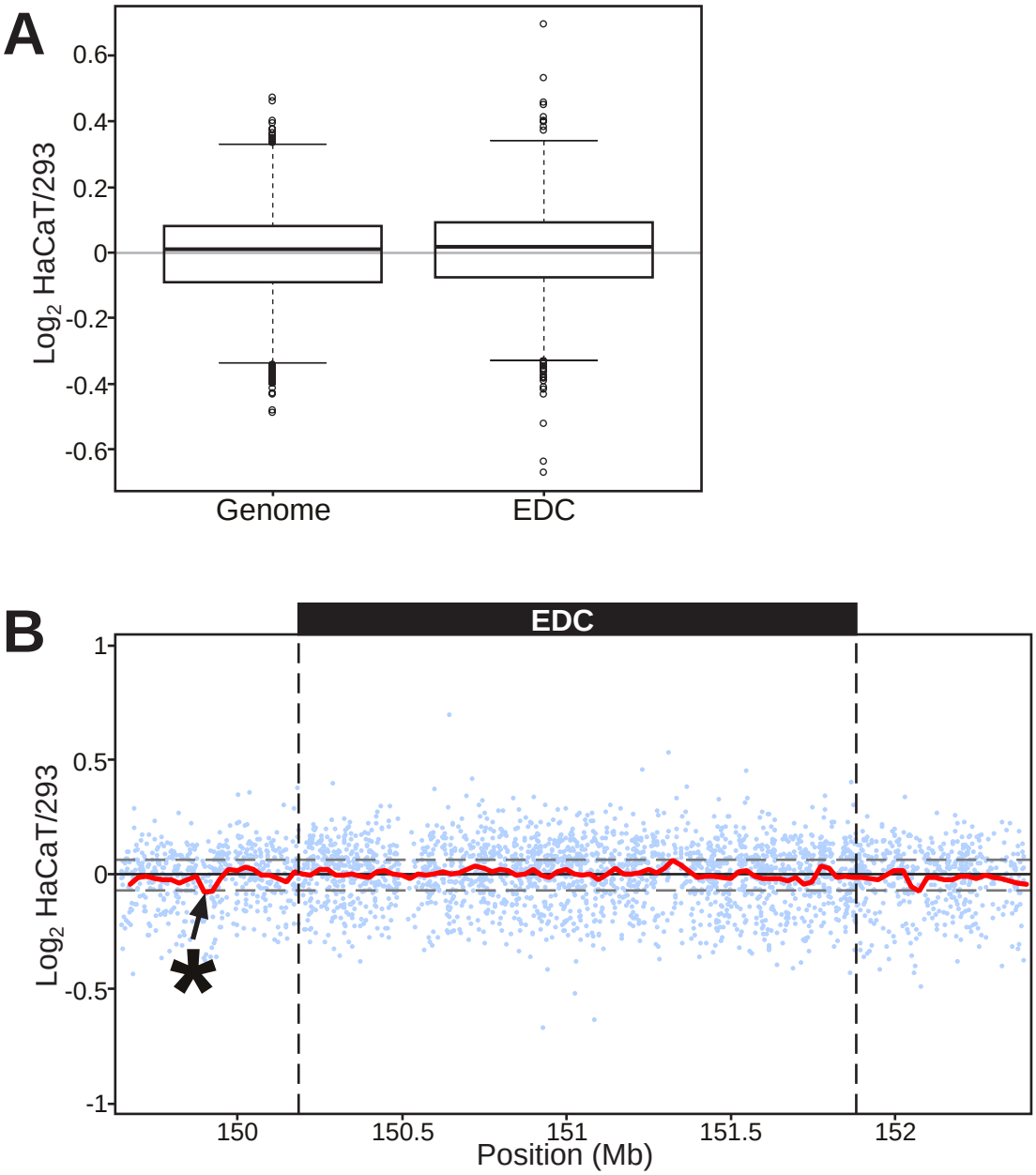


Figure 4.3 Array CGH between HaCaT and 293 Cells

A. Boxplot of $\log_2$ HaCaT/293 genomic DNA ratios from aCGH on a custom tiling array. The distribution of $\log_2$ values from the EDC (2145 oligos) was compared to the rest of the genome (2704 oligos). Their medians are 0.019 and 0.009 respectively and the two distributions are significantly different (KS-test). The median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles).

B. Scatter plot of $\log_2$ HaCaT/293 genomic DNA ratios from array CGH. The data in the region around the EDC was plotted against its chromosomal position (taken from NCBI Build 36). The mean $\log_2$ values of individual oligos are indicated by the blue dots. The red line shows the mean value across 50Kb intervals. The dashed grey lines indicate the 99.5% confidence intervals for a true difference in copy number for 57 oligos. The asterisk inicates a region in which the running mean lies outside of these intervals.

the $\log_2$ ratio between the channels should be equal to 0 for every probe and the observed variation is caused by the technical error in the experimental system. By randomly sampling values from this dataset I can determine the probability of observing the median $\log_2$ value seen in my experimental hybridisations for the EDC through error alone. I therefore generated 10,000 random samples of 2145 probes from the control data and calculated the median $\log_2$ value of each. 0.01% of the sample medians were greater than or equal to 0.019 (the experimentally observed median $\log_2$ value of the EDC). Therefore, it is possible, but unlikely, that this experimental result could be generated purely by random error. The observed $\log_2$ value is only slightly >0 and therefore is unlikely to represent a difference in copy number of the EDC. It is more likely that a difference in copy number of part of the locus, causes a smaller copy number difference to be observed when the whole of the locus is measured.

To examine this, I plotted the $\log_2$ HaCaT/293 ratios against genomic position (*Figure 4.3B*). I next calculated the mean $\log_2$ value of each 50Kb interval across the region (median number of probes, 57). The 50Kb mean value varied around 0 along the length of the EDC. To determine whether any the values represented a disruption to the EDC I again used my control data to estimate technical error. First I calculated the mean $\log_2$ value of every group of 57 probes from the control data and then calculated the $\log_2$ interval within which 99% of these values lay. This allowed me to assign 99.5% confidence intervals for a relative higher or lower copy number in HaCaTs vs 293s. Any 50Kb region with a $\log_2$ value outside of the confidence intervals probably represented a true difference in copy number above technical error. None of the 50Kb intervals across the EDC lay outside of the confidence estimates. However, one region lying outside of the EDC and towards the centromere has a –ve mean value that lies outside of the confidence intervals, suggesting a greater copy number in 293 cells. This region lies within the house keeping gene *SNX27* whose expression status was assessed by RT-PCR in the previous chapter (see, *Section 3.1*, p110). The 50Kb interval only contains 30 probes and therefore the technical error in measuring its relative copy number is likely to be greater than for a region containing 57 probes. Since the mean for the region lies just outside the –ve confidence intervals it is unlikely to represent a true difference in

copy number. Therefore, from this analysis and the FISH, I conclude that their have been no specific gains or losses of DNA in the EDC in HaCaT or 293 cells.

4.2 Nuclear Positioning of the EDC in Model Cell Lines

As for a number of other gene clusters, the position of the EDC within the nucleus has been reported to differ between its active and inactive states (Williams et al., 2002). It was found that in primary keratinocytes, the EDC is positioned outside of the Chromosome 1 territory and in lymphoblasts, where the EDC is inactive, it adopts a more peripheral or internal location. I used a similar methodology to assess whether the EDC was also differentially positioned within the nuclei of my model cell lines.

The probes used to examine the karyotypes of the cell lines are also suited to the study of the nuclear position of the EDC (*Figure 4.1*). Probe 1 is positioned within the *SPRR* cluster and my RT-PCR analysis suggests that this region is the part of the EDC which is differentially expressed between the cell lines (*Figure 3.1*, p112). The other two probes can be used as controls because they are positioned outside of the EDC. Probe 2 lies very close to the EDC and may be expected to follow the behaviour of the EDC. Probe 3 on the other hand is ~2.5Mb away towards the telomere and therefore would be expected to be positioned differently from the other two probes.

The probes were hybridised together with a chromosome 1 paint to methanol-acetic acid fixed nuclei of each of the three cell lines (*Figure 4.4A*). Images of 60 interphase nuclei were taken for each probe and cell line and analysed using a custom script used previously in a number of studies by our laboratory (see *Section 2.11.1*, p97 and (Chambeyron and Bickmore, 2004;Gilbert et al., 2004;Mahy et al., 2002a;Mahy et al., 2002b). The script measures the distance from the centre of each individual probe signal to the edge of its parent chromosome territory and were originally developed by P. Perry and N. Mahy. I modified this method slightly because HaCaT and 293 cells are both aneuploid for chromosome 1 and it was often

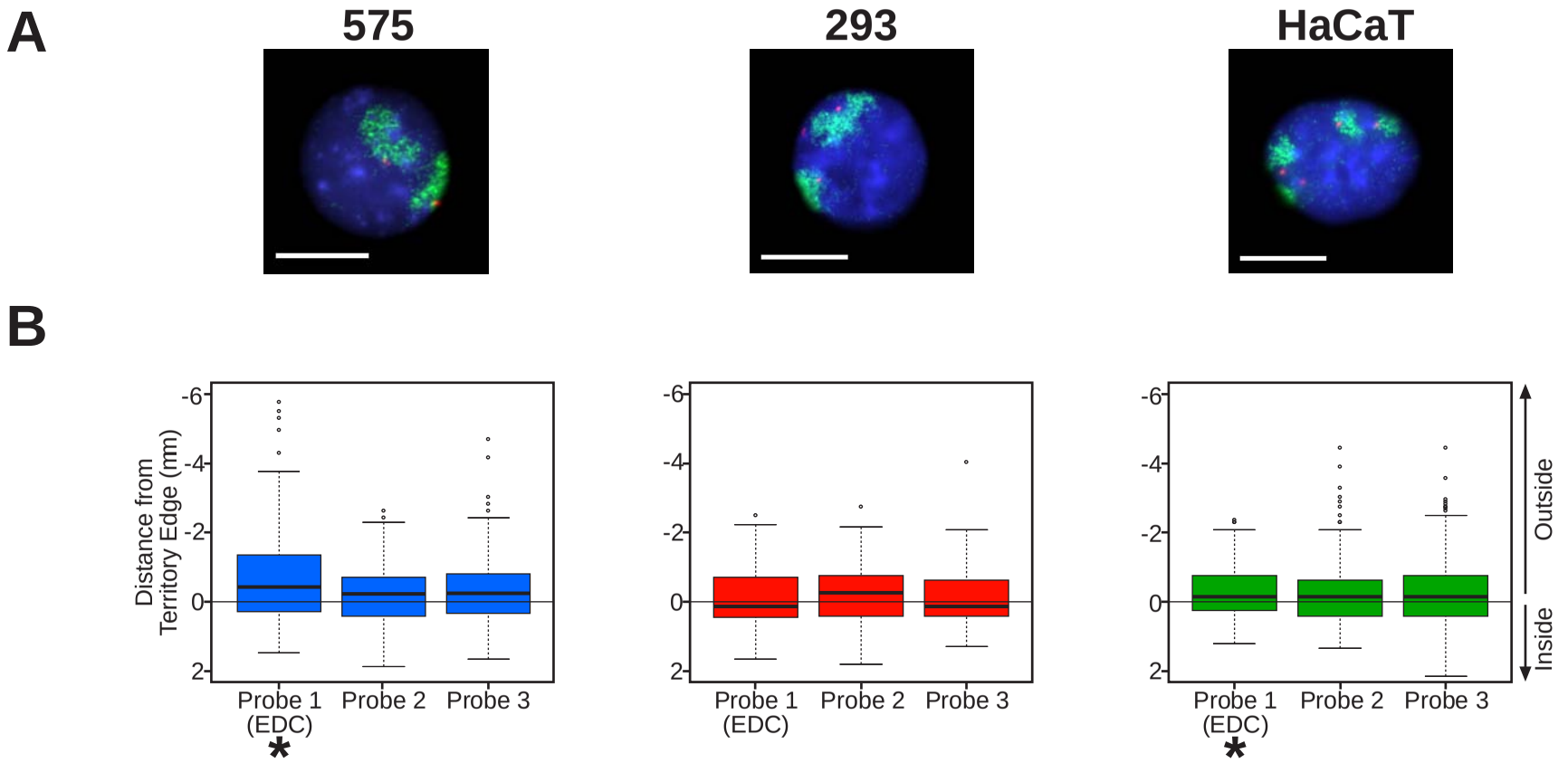


Figure 4.4 The Position of the EDC with respect to the Chromosome 1 Territory

A. Individual nuclei from each of the three cell types are shown after FISH with the chromosome 1 territory (green) and probe 1 (the EDC, red). DNA is counterstained with DAPI (blue). The Scale bars indicate 10mm.

B. Boxplots showing the distribution of distances from the probe signal to the chromosome 1 territory edge for the three cell lines. Probe 1 is the EDC and Probes 2 and 3 lie outside of the EDC (see *Figure 4.1*). Negative distances represent positions outside the chromosome territory and positive distances inside the territory. For each boxplot the median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles). Asterisks inidicate distributions that are significantly different from the others in a particular cell line.

difficult to unambiguously identify which probe signal belonged to which chromosome territory. Therefore, I measured the distance from each BAC to the nearest chromosome territory edge.

The relative distributions of probe to territory edge distances for each cell line were then examined using boxplots (*Figure 4.4B*). In 575 cells all of the probes occupied a similar median position, outside, but close to the edge of the chromosome 1 territory (median positions: -0.423, -0.218 and -0.237μm respectively). Probe 1 occupied a slightly, but significantly, more external position in relation to the other probes and no statistical difference was found in the locations of probes 2 and 3 (KS-tests). All probes were also positioned close to the territory edge in 293 cells (median positions: 0.134, -0.256 and 0.134μm respectively), but no significant differences were found in the overall distributions (KS-tests). The median position of all three probes in HaCaT cells was identical (-0.134μm). Probes 2 and 3 occupied a wider range of positions than probe 1. However, I only found statistical differences between probes 1 and 3, the p-value of a KS-test between probes 1 and 2 was close to significance (0.069) and the positions of probes 2 and 3 were not significantly different.

These results suggest that the EDC (probe 1) occupies a similar position in all three of the cell lines, outside but close to the edge of the chromosome 1 territory. The two control probes also occupy similar positions in each cell line, which suggests that the entire region extending out from the EDC and into chromosome band 1q22 may be positioned close to the edge of the chromosome 1 territory. The entire region is known to be gene-rich and these results are consistent with previous reports that, in general, gene-rich areas are positioned on the edge or outside of chromosome territories irrespective of their gene expression status (Gilbert et al., 2004;Mahy et al., 2002a;Mahy et al., 2002b). Next I directly compared the position of the EDC between the cells lines and found probe 1, but not probes 2 and 3, is differently positioned in each cell type (KS-tests, p-values < 0.05). Therefore, although the EDC is positioned close to the edge of the chromosome 1 territory in each cell line, it actually occupies a slightly different position in each cell type. However, contrary to a previous report, I find that in HaCaT

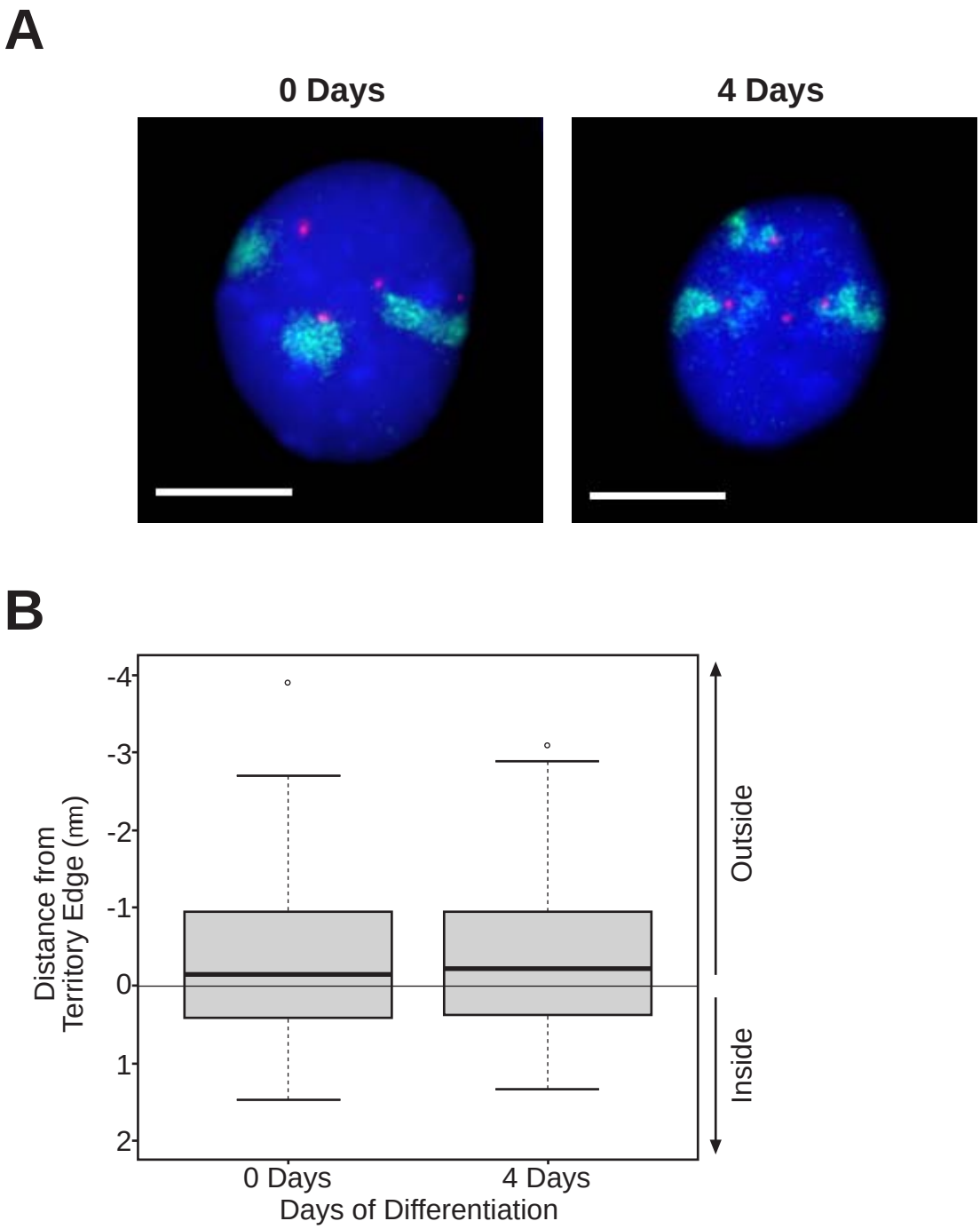


Figure 4.5 Nuclear Position of the EDC during HaCaT Differentiation

A. FISH images of individual HaCaT nuclei before (0 days) and after 4 days of differentiation. The chromosome 1 paint is in green, probe 1 (see, *Figure 4.1*) is in red and DNA is counterstained with DAPI (blue). The scale bar indicates 10mm.

B. Boxplot showing the distribution of distances from probe signal to the edge of the chromosome 1 territory for the two cell populations. Negative distances lie outside the territory and positive distances inside. The median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles). n= 149 for 0 days and 159 for 4 days.

keratinocytes the EDC does not occupy a more external position than it does in lymphoblastoid cells (Williams et al., 2002).

4.3 Differentiation Does Not Alter the Position of the EDC in HaCaT Cells

In the previous chapter, I characterised the transcriptional changes that are induced upon *in vitro* differentiation of HaCaT cells (*Section 3.6*, p126). I found that after 4 days of differentiation the cells upregulate the *SPRR* cluster and BAC probe 1 lies within this region. I therefore, used FISH to determine the position of probe 1 relative to the chromosome 1 territory edge in either undifferentiated cells or cells that had been differentiated for 4 days (*Figure 4.5A*). A boxplot comparing the distributions of distances from the territory edge, before and after differentiation, showed that the two were not significantly different (*Figure 4.5B*, median positions: 0.134μm and 0.218μm respectively, KS-test, p-value=0.990). Therefore, I can conclude that *in vitro* differentiation of HaCaT cells does not significantly alter the nuclear position of the *SPRR* cluster, despite inducing activation of genes within that region.

4.4 A Different Profile of Interphase Chromatin Compaction at the EDC in HaCaT Cells

As well as measuring the nuclear position of a locus, FISH can be used to assess chromatin compaction. There is a linear relationship between the genomic separation between two FISH probes and the mean-squared interphase distance between them in the nucleus (mean d^2, van den Engh et al., 1992). This is based on a random walk model for the path the chromatin fibre takes between the two probes and appears to hold true for genomic separations between ~50Kb and 2Mb (Sachs et al., 1995;Yokota et al., 1995). This method was then applied to measure regional differences in the compaction of chromatin (Yokota et al., 1997). Our group has also

used this assay to find similar differences in regional compaction of chromatin in human lymphoblastoid cells (Gilbert et al., 2004) and to show that both the murine *HoxB* and *HoxD* loci cytologically decondense upon induction of differentiation in both mouse embryonic stem cells and the embryo (Chambeyron and Bickmore, 2004;Chambeyron et al., 2005;Morey et al., 2007). I decided to use this assay to examine chromatin compaction across the EDC in my model cell lines. I selected 3 pairs of Fosmid probes (average size ~40Kb) separated from each other by 500Kb that spanned the EDC (*Figure 4.6A*). Two further pairs of probes with a similar separation were selected to cover the regions just outside of the EDC and to allow me to examine how the compaction in the EDC differed from its surroundings.

Firstly however, I tested a control locus to discover if any systematic differences in chromatin compaction existed between the cell lines. Statistical tests of my expression microarray data showed no difference in expression between the cell lines at the *β-globin* locus (KS-tests). Therefore, I selected a pair of fosmid probes separated by 500Kb that spanned this locus and performed FISH with them on methanol-acetic acid fixed preparations of the three cell lines (*Figure 4.6B* and *C*). I took pictures of at least 100 nuclei from each cell type and calculated the distance separating the signals of each probe pair using a custom script based on one used previously in our laboratory but modified to facilitate faster analysis of large numbers of nuclei (*Section 2.11.2*, p98 and (Chambeyron and Bickmore, 2004). The distribution of observed distances for each cell line conformed to a Rayleigh distribution (*ie* the Standard Deviation/Mean ~0.52 and the Median/Mean ~0.94), confirming previous reports that at these genomic distances the chromatin fibre between the two probes follows a random walk model in the nucleus (Sachs et al., 1995;Yokota et al., 1995). The mean d^2 for each of cell line was normalised by the genomic distance separating the probes (mean d^2 per Mb) and plotted to compare the compaction for each cell line (*Figure 4.7A*).

There was no significant difference in the packing of the *β-globin* region between the control cell lines (293 and 575) but the region was significantly more compact in HaCaT cells (KS-tests). This might be explained by a difference in nuclear size between the cell lines. However, although HaCaT nuclei were generally smaller than those of 293 cells, they were larger than 575 cell nuclei (*Figure 4.7B*)

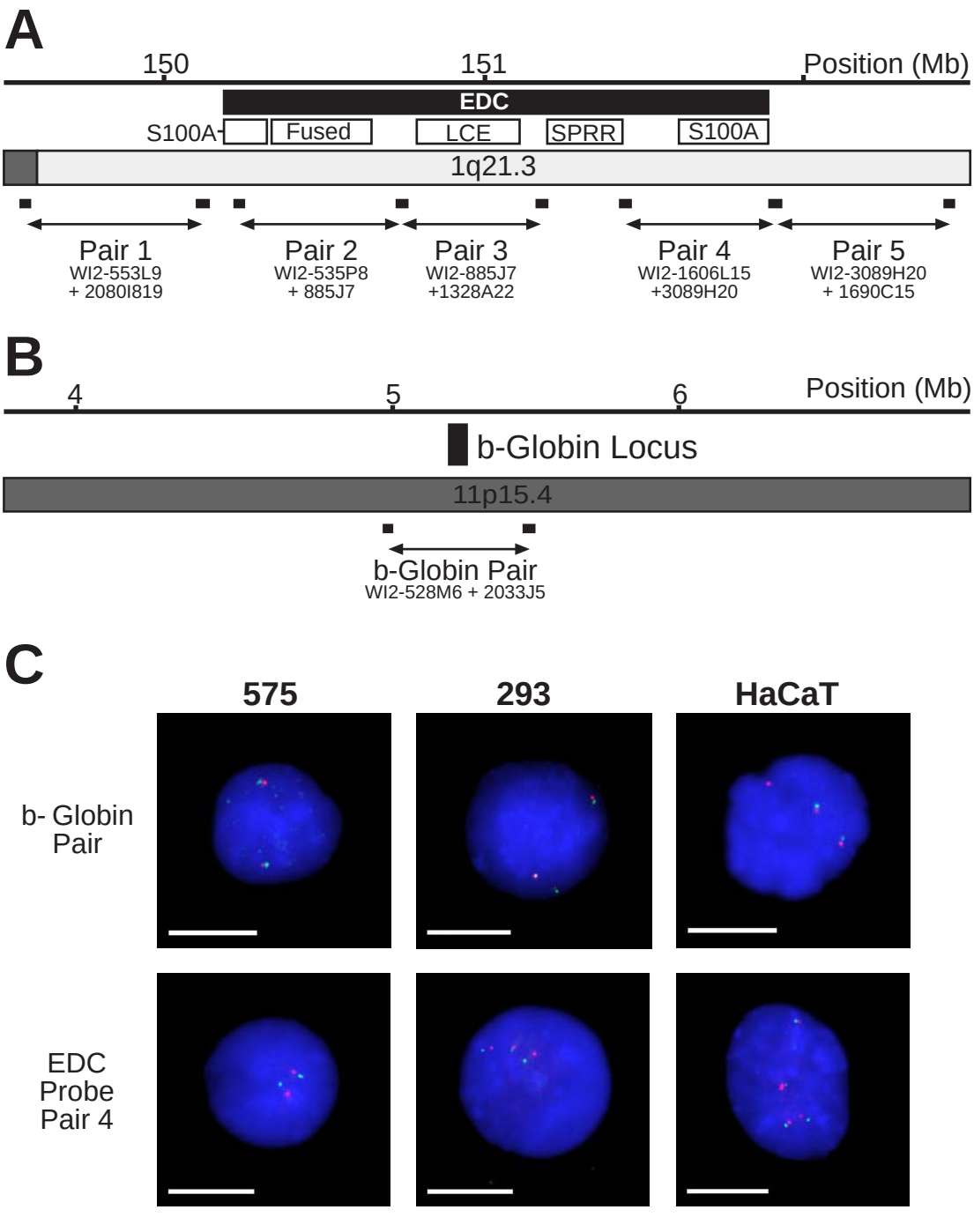


Figure 4.6 Measuring Interphase Condensation at the EDC

A. Locations of probe pairs used to measure interphase condensationof the EDC. The genomic positions are taken from the UCSC browser (March 2006, NCBI v36 Build). Fosmid probes used in each pair are shown to scale along with their universal names (for the locations of each fosmid see *Section 2.7.9* p80.).

B. As panel A, but for the b-Globin Locus on human chromsome 11.

C. FISH images of nuclei using either the b-Globin Probe pair or EDC Probe Pair 4 from each cell line. The individual fosmids are shown in red and green and DNA is counterstained with DAPI (Blue). The scale bars indicate 10mm.

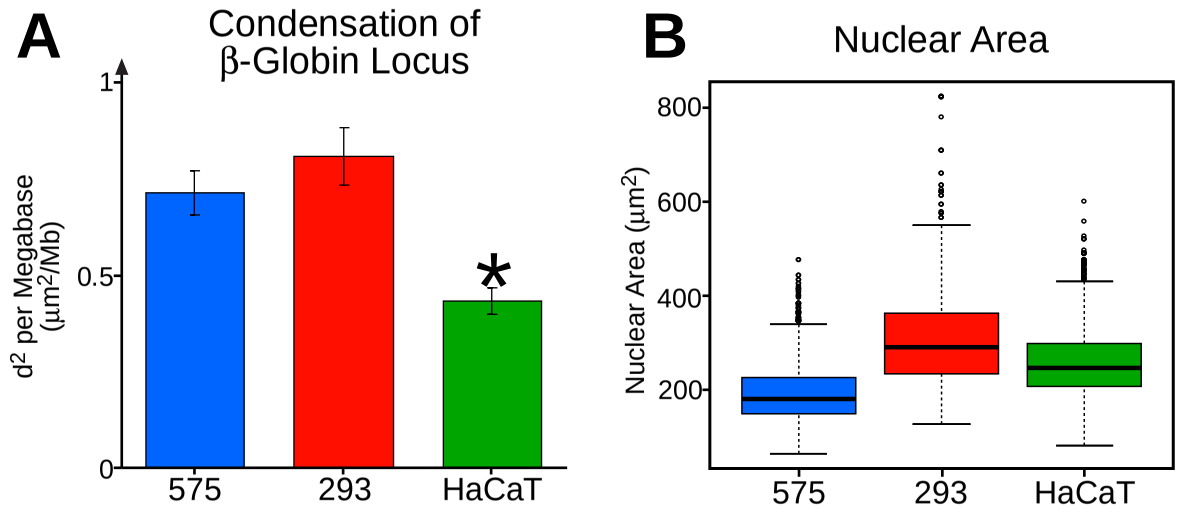


Figure 4.7 Overall Interphase Condensation Differs in HaCaT Cells

A. The interphase separation of the β-Globin Probe Pair (*Figure 4.6B*). Mean squared probe separation per Mb (d^2) is plotted along with the standard error in the mean (n>200 for all cell lines). The mean probe separation of the the locus in HaCaT cells is significantly different from the other two cell lines (KS-tests, p values < 0.001).

B. Boxplot of observed nuclear area calculated from FISH images of each of the three cell lines. The areas were calculated assuming a flattened, circular nucleus and using images taken from both the β-Globin pair and the EDC pairs. The median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles).

meaning that the greater compaction observed at the *β-globin* locus in HaCaT cells is not due to a smaller nucleus. However, the three cell lines all have different numbers of chromosomes (575 cells are diploid, 293 cells are roughly triploid and HaCaT cells roughly tetraploid). Therefore, the three cell types may have a different ratio of genome size to nuclear area, at least in methanol-acetic acid preparations, causing the chromatin to be more compact in this assay. I calculated the nuclear area per Megabase of DNA for each cell line in order to determine if this was the case: in 575 cells there are $0.028\mu m^2$/Mb, in 293 cells, $0.030\mu m^2$/Mb and, in HaCaTs $0.019\mu m^2$/Mb (assumes that the nuclei have flattened to circles and that the human diploid genome is 6391Mb in size). Therefore, HaCaTs do have a lower DNA to nuclear area ratio than 575 or 293 and this may explain the difference in compaction that I observe. This may not solely be an artefact of the cell lines, as studies of nuclear size in both normal and psoriatic skin have found that, during normal epidermal differentiation, the nucleus decreases in size (Tsuji and Cox, 1977).

As expected from analysis of the *β-globin* region, the EDC was also more compact in HaCaT cells than in 575 and 293 cells (*Figure 4.8*). However, the chromatin compaction profile across the region differed between the cell lines. In the two control cell lines the interphase separation of probe pair 4 was significantly greater than any of the other probe sets (KS-tests). In HaCaT cells, the profile of chromatin compaction was more 'flat' than that of the two control cell lines. Furthermore, Probe pairs 4 and 5 were significantly less compact than pairs 1, 2 and 3 in these cells. Probe pair 2 was also more condensed than pairs 1 and 3 but only the compaction of pairs 2 and 3 were significantly different (KS-tests, p-values: 0.063 for probes 1 and 2, 0.036 for probes 2 and 3).

The profile of chromatin structure at the active EDC is different to that of the inactive EDC. The larger *S100A* gene cluster (probe pair 4, *Figure 4.6A*) is constitutively decondensed compared to the rest of the region. However, when the EDC is active this decondensed structure extends out of the EDC towards the telomere (*ie* into the region covered by probe pair 5). The rest of the inactive EDC has a more compact structure that extends outwards towards the centromere. However, when active the region covering *S100A10* and *A11* along with the *fused* cluster appears to adopt an even more compact structure.

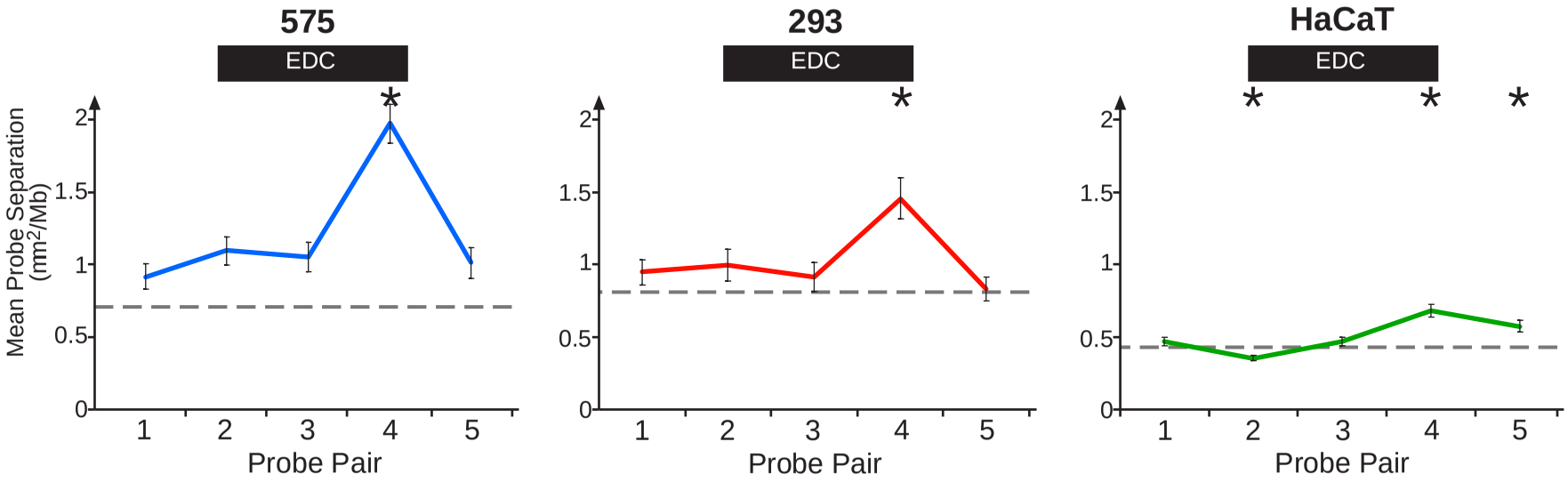


Figure 4.8 Interphase Condensation of the EDC

The mean d^2/Megabase (mm^2/Mb) +/- the standard error in the mean was plotted for each cell type. For probe pair locations see, *Figure 4.6*. The dashed grey line on each plot shows the mean d^2/Megabase of the b-Globin locus for each cell type (*Figure 4.7*). Observations indicated by an asterisks are significantly different from others in the same cell type at the 5% level (KS-test). For all cell lines and probes the sample size, n > 100 and the distributions of observed distances conformed to a Rayleigh distribution.

4.5 The Correlation between Transcription and Interphase Chromatin Organisation at the EDC

I next wanted to investigate the relationship between the transcriptional status of the EDC in my cell lines and their interphase chromatin compaction. I first identified which oligonucleotides on the human genome wide expression arrays lay within the regions covered by each EDC probe pair (probe positions from, van den Ijssel et al., 2005, see *Chapter 3*). I then calculated a median $\log_2$ cell/reference cDNA value for each region and cell type and plotted them to compare the expression profiles of the three cell lines (*Figure 4.9A*). The two control cell lines both had similar expression profiles but HaCaT cells had higher median $\log_2$ values in both regions 2 and 4 (corresponding to the *fused* cluster and the two *S100A* clusters).

We might expect that a higher expression level would correlate with a lower level of chromatin compaction. However, this does not appear to be the case, the highest expression level is at region 2 in HaCaT cells but this is in fact the most compact of the five regions in this cell line (*Figure 4.8*). Furthermore, expression levels in both 575 and 293 cells are lowest in region 4, the least condensed region in both of these cell lines. However, as noted in Chapter 3, cell/reference ratios are difficult to interpret with regard to the actual expression level of a particular gene or region. For, example a low ratio could arise for a very active gene in a particular cell line if it was also active in the cell lines from which the reference RNA is made. Therefore, I calculated relative expression ratios between HaCaTs and the two control cell lines for each region. I first calculated a HaCaT/575 or 293 $\log_2$ ratio for each probe before deriving the median HaCaT/Control $\log_2$ value for each FISH region. In order to compare these relative expression levels to the chromatin compaction of each region I then had to calculate relative compaction values for each region (ie the mean d^2/Mb ratio between HaCaTs and control cells). I then plotted the relative expression level against the mean d^2/Mb ratio for both HaCaT/575 and HaCaT/293 (*Figure 4.9B*). The two were negatively correlated, both when the whole dataset was considered (Pearson R^2 value = -0.681), and when the HaCaT/575 and HaCaT/293 values were considered separately (R^2= -0.751 and, -0.712 respectively).

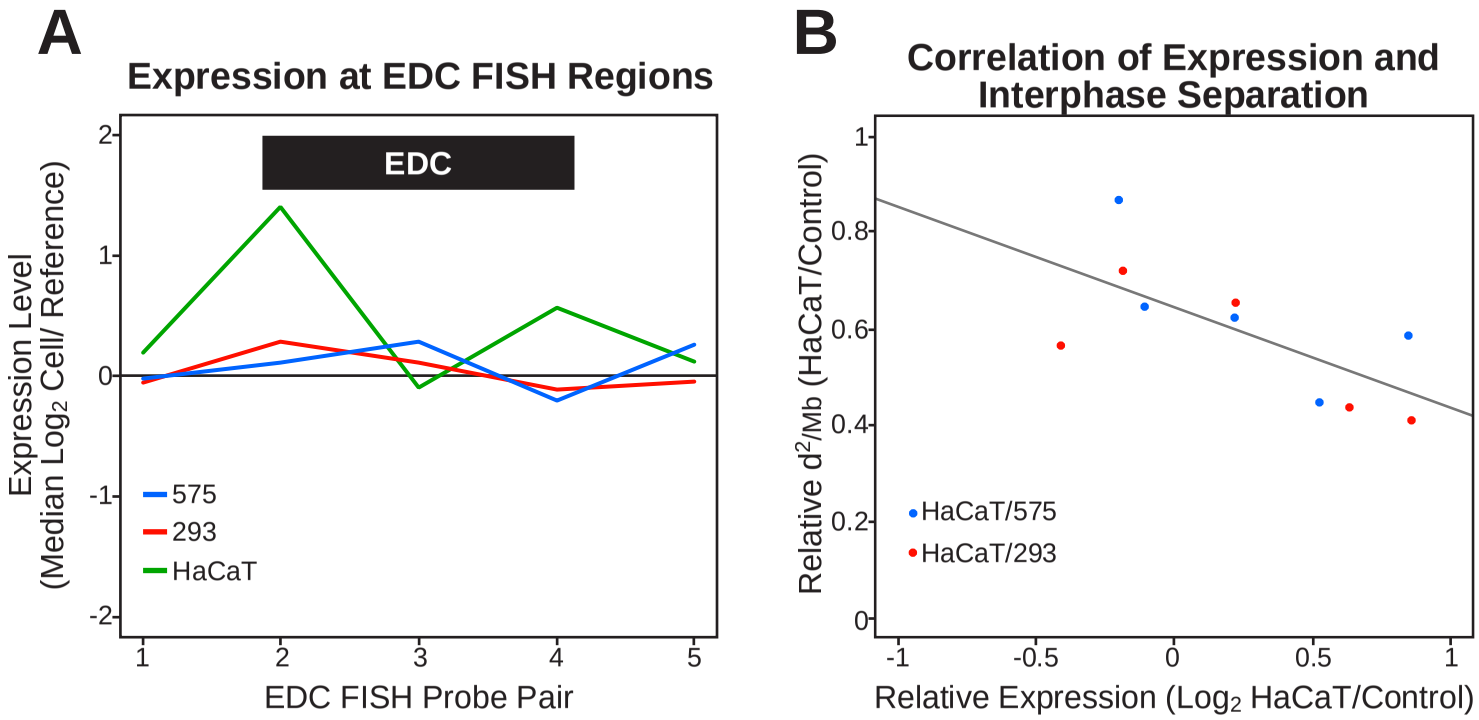


Figure 4.9 Comparison of Chromatin Compaction and Expression Levels at the EDC

A. Median $\log_2$ Cell/Reference values of each of the EDC FISH regions (see *Figure 4.6*) were calculated from genome wide expression data (see *Chapter 3*). 575 celss are in blue, 293 in red and HaCaT in green.

B. Median $\log_2$ HaCaT/Control relative expression levels plotted against HaCaT/Control mean d^2 per Mb value. A line of best fit drawn was drawn (R^2, Pearson's = -0.681). Values in blue are HaCaT/575 ratios and in red are HaCaT/293.

Therefore, at the EDC for a 500Kb genomic separation, surprisingly, there is a negative relationship between interphase chromatin compaction and expression levels. This is contrary to what we might expect since previous studies have correlated gene activity to a cytological decondensation of the *Hox* loci (Chambeyron and Bickmore, 2004;Chambeyron et al., 2005;Morey et al., 2007). However, these studies examined compaction across a smaller genomic distance, for example 90Kb for the *HoxB* locus. It is therefore possible that the same analysis on the EDC using probes separated by smaller genomic distances would also show a positive correlation.

4.7 Discussion

Here I have examined the interphase organisation or higher order chromatin structure of the EDC in cultured cell lines. Previously it has been reported that the active EDC is positioned outside of the chromosome 1 territory (Williams et al., 2002). I have shown that this also the case in cultured HaCaT cells but contrary to this report I find that although the inactive EDC in my control cell lines occupies a different position, it is also positioned externally. I have also shown that two flanking regions occupy similar locations in all these cell types and that the induction of differentiation of HaCaT cells does not significantly alter the position of part of the EDC. I have then gone on to analyse the interphase organisation of the EDC in more detail by comparing the separation of pairs of probes in the interphase nucleus. I show that, in all three cell lines the larger *S100A* gene cluster at the telomeric end of the EDC is decondensed relative to the rest of the locus. However, when the EDC is active the region around the other *S100A* cluster and the *fused* genes is more condensed and a region lying outside the EDC is decondensed. In addition, it appears that in general the interphase chromatin of HaCaT cells is more compact at this level than the other cell types, something that may also be the case in epidermal cells *in vivo* (Tsuji and Cox, 1977). I attempted to correlate the relative expression level of different parts of the EDC to the levels of interphase compaction. This suggests that, the higher expression levels observed at some parts of the EDC in HaCaT cells actually correlate with a relatively more compact chromatin structure, at least over large

genomic distances. Finally, I have also examined the karyotypes of each of the three cell lines and despite the presence of abnormal derivatives in both 293 and HaCaT cells, no disruption of the EDC had occurred due to chromosomal rearrangements. Next I wanted to examine how other levels of chromatin structure related to the interphase chromatin organisation of the EDC.

Chapter 5: Primary and Secondary Chromatin Structure Analysis at the EDC

Having studied the interphase organisation of the EDC, I next wanted to investigate how interphase compaction is related to lower orders of chromatin structure at the EDC. Recently we have begun to understand how primary chromatin structure (predominantly in the form of histone modifications) relates to transcription across the genome from a large number of Chromatin Immunoprecipitation (ChIP) Studies (Barski et al., 2007;Heintzman et al., 2007;Koch et al., 2007;Roh et al., 2005;Roh et al., 2006). A subset of histone modifications mark the promoters of genes but in some regions wider domains of modifications are seen, for example *Hox* loci and heterochromatin (Bernstein et al., 2005;Martens et al., 2005). However, we do not yet know if, and how, these modifications influence higher levels of chromatin structure. This is partly because few studies have examined the secondary chromatin structure of the genome. Sucrose gradient sedimentation has been used to examine the structure of both mouse minor satellite and the chicken *β-globin* locus (Caplan et al., 1987;Fisher and Felsenfeld, 1986;Gilbert and Allan, 2001;Kimura et al., 1983). More recently this type of biophysical analysis was combined with microarrays to examine chromatin fibre structure across the genome in lymphoblastoid cells (Gilbert et al., 2004). However, this analysis was performed at a low resolution and so could only be related to tertiary structure and not primary chromatin structure. Therefore, I have assayed chromatin fibre structure across the EDC to a much higher resolution using tiling oligonucleotide microarrays and related secondary chromatin structure to the tertiary organisation of the locus described in *Chapter 4*. By performing ChIP for a number of different histone modifications, I have then examined whether the distribution of histone modifications at the locus can be related to the secondary chromatin organisation of the region. Finally, a recent study has described a general depletion of nucleosomes from the region around active promoters in the human genome (Koch et al., 2007). However, the methodology used was likely to create

artefacts in the analysis of histone occupancy, therefore I have examined this using a more direct method.

5.1 HaCaT and 293 Cells Have Similar Bulk Chromatin Structures

My analysis of interphase chromatin compaction in the previous chapter suggested that, at the tertiary level, HaCaT chromatin may be more compact than that of 293 or 575 cells. I therefore, wanted to examine whether the cell lines exhibited the same differences in bulk chromatin structure at the level of secondary chromatin structure. For these experiments, I decided I would only use 293 cells as a control because, like HaCaTs, they grow in adherent cultures, whereas 575 cells grow in suspension.

There are a number of different techniques that can be used to assay secondary chromatin structure, however, many of them do not easily allow the structures of two cell lines to be compared. One method that has successfully been used to compare the bulk chromatin structure of different cell lines is sucrose gradient sedimentation (Gilbert et al., 2007). In principal, if two chromatin fibres from two different cell lines have the same mass and the same overall conformation, they will sediment similarly on a sucrose gradient. However, if the two have the same mass and a different conformation they will sediment differently. For example, if bulk chromatin from HaCaT cells had the same unit mass as chromatin from 293 cells but formed a more compact secondary fibre, we might expect that it would move through the gradient more rapidly.

However, before performing this analysis it is important to establish that bulk chromatin from the two cell lines does indeed have the same unit mass. Two DNA molecules of the same length obviously have the same mass, so differences in the unit mass in chromatin will be caused by differences in the mass of protein associated with the DNA. The repeat length is the length of DNA per nucleosome in the cell and is known to vary between cell lines (Allan et al., 1984;Stein and Mitchell, 1988;Weintraub and Groudine, 1976). If a cell line has a shorter repeat length this means that its chromatin will have a greater number of nucleosomes in a

given length of DNA and so a higher unit mass. Therefore, I analysed repeat length by digesting nuclei from HaCaT and 293 cells with micrococcal nuclease (MNase). After the enzyme was added to the nuclei, aliquots were removed along an exponential timecourse, DNA purified from them and resolved on an agarose gel (*Figure 5.1A*).

To calculate the repeat length of bulk chromatin, I calculated the DNA size of each visible nucleosomal band from the gel images. For each timepoint I calculated the repeat length from the gradient of a graph of nucleosome number (ie mono-, di-, tri, tetra, penta, etc) against DNA fragment size using linear regression. When repeat length was plotted against time, the profiles of the two cell lines appeared very similar which suggests they have the same repeat lengths (*Figure 5.1B*). The repeat length *in-vivo* could be calculated by extrapolating the graph back to time zero but this is difficult because as we approach time zero the graphs become asymptotic.

Another method, of examining bulk chromatin structure is to analyse its digestion properties. Chromatin with different compaction properties digests differently with MNase (Gilbert et al., 2007). Therefore, to further compare the properties of bulk chromatin of 293 and HaCaT cells, I calculated the digestion rate of high molecular weight chromatin using a previously published method (Gilbert et al., 2007). By defining a boundary on the gel images as midway between the penta- and hexa- nucleosome bands, I measured the total intensities of high and low molecular weight DNA at each timepoint (*Figure 5.1A*). Then I calculated the percentage of high molecular weight DNA at each timepoint and normalised the profiles by defining the 1min timepoint as 100%. The digestion profiles of the two cell lines were very similar and the half lives of 293 and HaCaT cells were estimated as 6.2 and 6.7mins respectively graph (*Figure 5.1C*). Therefore, as well as having a similar repeat length, bulk chromatin from HaCaT and 293 cells digests at a similar rate.

Having established that chromatin from the two cell lines was likely to have the same unit mass because of their similar repeat lengths, I then analysed the sedimentation properties of bulk chromatin. To prepare chromatin for sucrose gradient sedimentation, nuclei were isolated from the two cell lines and resuspended to an equal DNA concentration. Small amounts of MNase nuclease were then used

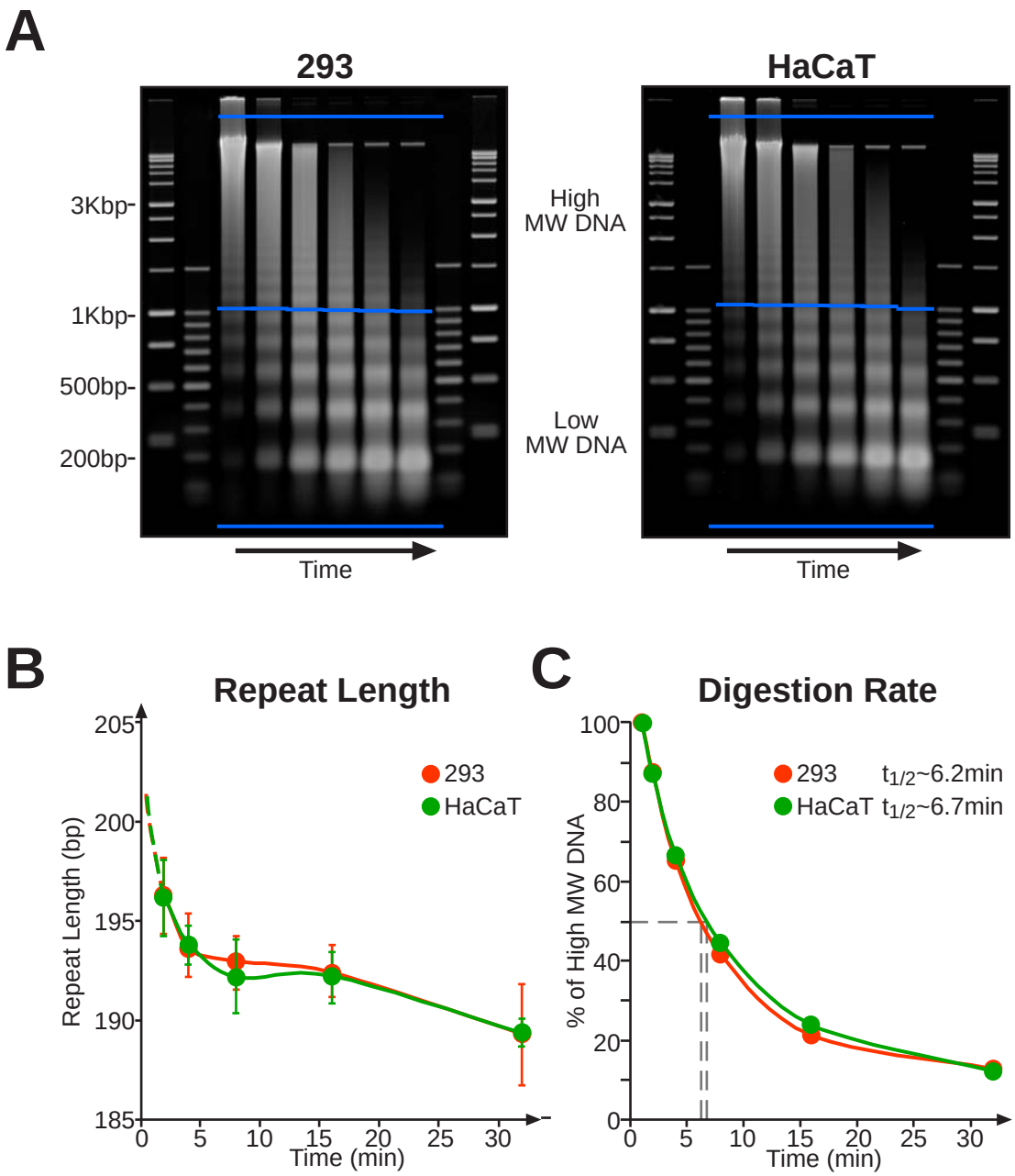


Figure 5.1 Comparison of MNase Digestion properties of HaCaTs and 293s

A. Images of the digestion time course analysis gels. The boundry between high and low molecular weight DNA for digestion rate analysis was set at half way between the penta- and hexa- nucleosome fragments. High molecular weight DNA lies between the upper two blue lines and low molecular weight between the lower two.

B. Graph of repeat length over time was plotted for HaCaT and 293 cells. Repeat length at each timepoint was calculated from the gradient of a graph of nucleosome fragment number (ie mono-, di-, tri-, tetra-, etc) against DNA length for each time point. Observations represent the mean of three experiments and the error bars represent the standard error in the mean.

C. Graph of digestion rate versus time for the two cell lines. The percentage of high molecular weight DNA was calculated by dividing the intensity of high molecular weight DNA on the gel by the intensity of total DNA (High + Low MW). The percentages were expressed by normalising time= 1min as 100%. The half life for the digestion of high molecular weight DNA is indicated by the dashed grey lines.

to digest the chromatin and the reaction stopped by the addition of EDTA. Digested nuclei were spun down gently, resuspended in buffer and incubated overnight on ice to release soluble chromatin. The next morning the nuclei were again spun down and the supernatant containing chromatin loaded onto a 6-40% isokinetic sucrose gradient and spun for 3.5 hours at 41,000rpm in a Sorvall SW41 rotor. Fractions were pumped off the gradient by upward displacement through a UV monitor and then collected. An aliquot of soluble chromatin was also collected from the supernatant to be used as an input fraction. Analysis of the DNA present in the pellet showed that it largely contains undigested genomic DNA (data not shown).

First I analysed traces of the chromatin from each gradient generated from the reading of the UV monitor and found that the amplitude of the chromatin peak was lower in HaCaT cells than 293s (*Figure 5.2A*). The position of the HaCaT peak fraction was also slightly higher up the gradient. Similar observations were made from a further independent gradient from each cell line. Given that we know the two cell lines digest at a similar rate, the difference in amplitude is most simply explained by a differential chromatin release from the two cell lines. It is not clear why the peak fraction should be different.

To further analyse the protein composition of bulk chromatin, I prepared protein from selected fractions using Trichloroacetic Acid (TCA) precipitation and analysed them on SDS PAGE gels (*Figure 5.2B*). As well as the core histones, the chromatin fractions all contained bands corresponding to the linker histones (*Figure 5.2B*, asterisks). Linker histones are much less tightly associated with the chromatin fibre than the core histones and if they have remained bound to chromatin during sedimentation, it suggests that the conformation of the chromatin was not disrupted during isolation and sedimentation (Allan et al., 1981). Furthermore, the fractions from 293 cells have also been examined by western blot in a separate study to show that the proteins LEDGF and HP-1α remain chromatin bound (H. Sutherland, personal communication). There appeared to be a difference in either the abundance of the different linker histone subtypes or their modification levels because in 293 cells two linker histone bands of roughly equal intensity were seen on the gel but in HaCaT cells, the lower band appeared less intense than the upper band (*Figure 5.2B*, asterisks). To determine if this is truly the case, the linker histone compositions

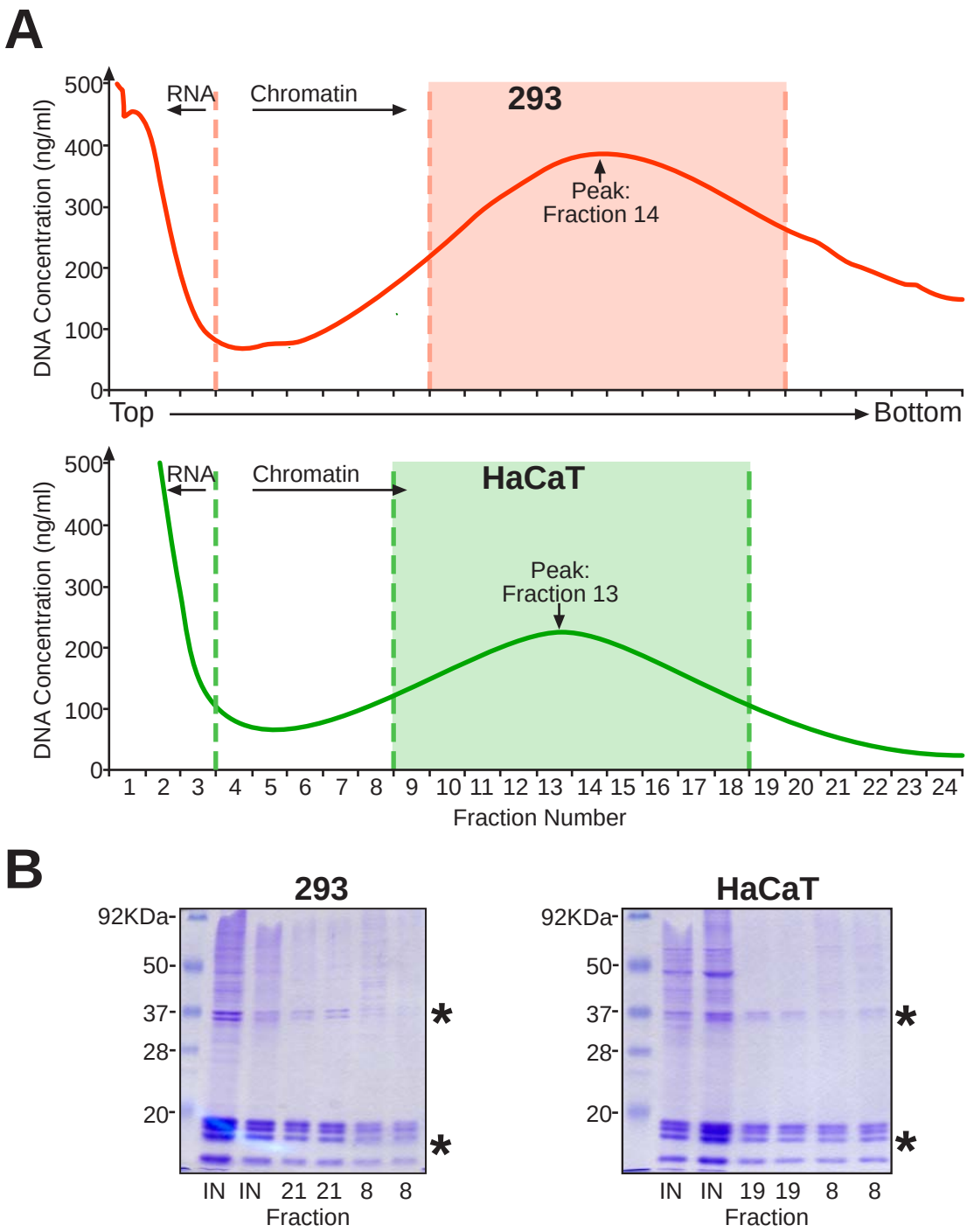


Figure 5.2 Analysis of Bulk Chromatin by Sucrose Gradient Sedimentation

A. Chromatin (Absorbance) traces of sucrose gradients. Shown below the graphs are the fraction numbers and the highlighted areas in each graph indicate the fractions from which DNA was isolated. The top fractions are believed to contain RNA and the traces show gradients processed in parallel. The peak fraction of each gradient is indicated also.

B. Coomassie stained gels of protein isolated from sucrose gradients. Fractions of both fast and slow sedimenting chromatin (ie Top and Bottom) were isolated from gradients. The asterisks indicates the core (lower) and linker (upper) histones which are present in every fraction. The two repeats of each fraction are isolated from independent gradients.

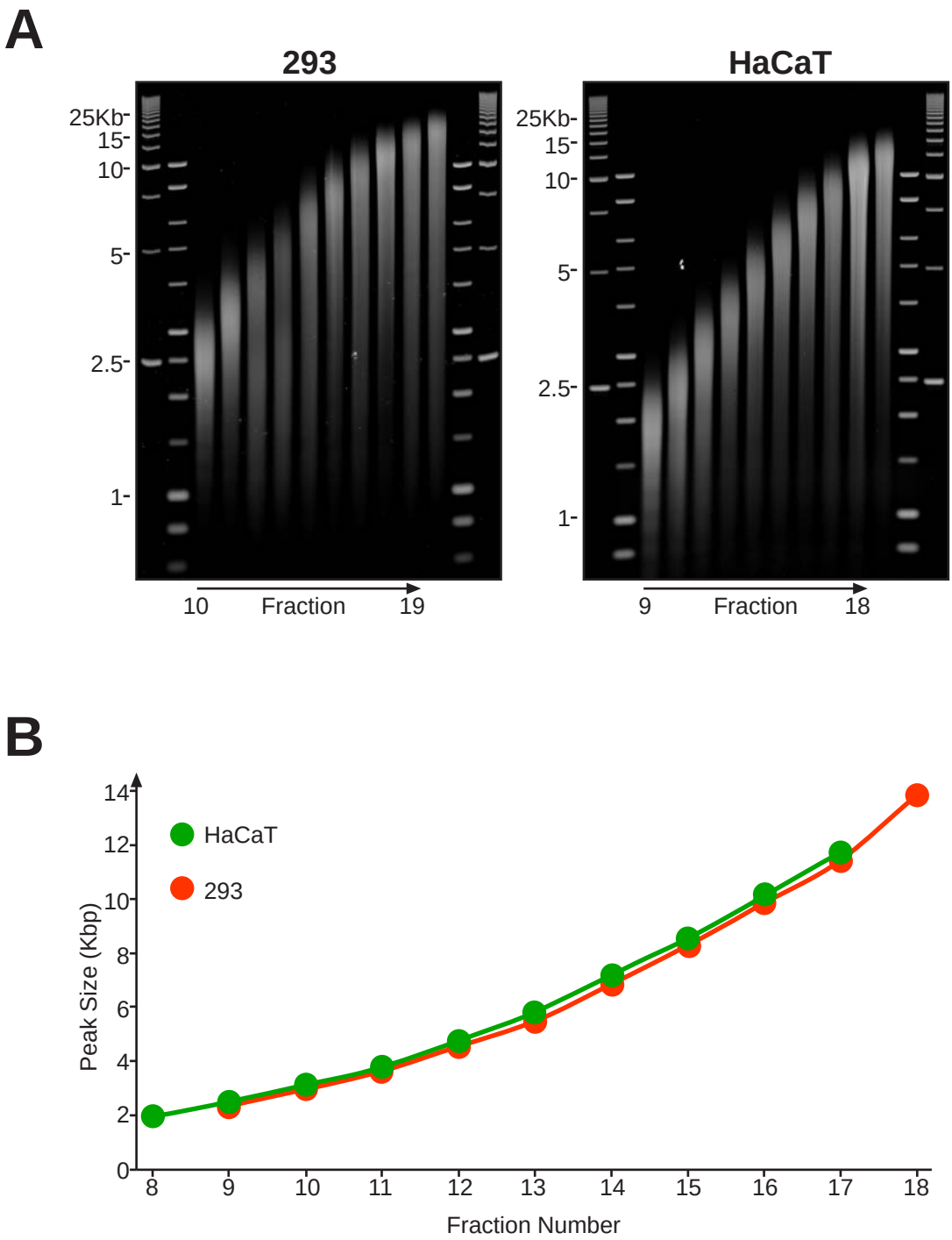


Figure 5.3 Analysis of Bulk Chromatin Sedimentation Rate

A. DNA isolated from sucrose gradients resolved on TPE Agarose gels. Fractions 10 - 19 were isolated from 293 cell gradients and 9-18 from HaCaT gradients (*Figure 5.2*).

B. Plot of peak DNA size against fraction number. The DNA size of the intensity peak for each lane was calculated from gel images. Each data point represents the mean of two independent gradients.

would have to be analysed using specific antibodies or mass spectrometry.

Next I isolated DNA from the fractions corresponding the chromatin peak of the two cell lines (*Figure 5.2A*, shaded area). The DNA was then separated on Tris-Phosphate (TPE) buffered agarose gels to examine the molecular weight of DNA found in each fraction (*Figure 5.3A*). TPE was used as a buffer because it gives better resolution of high molecular weight DNA fragments (N. Gilbert, PhD thesis, University of Edinburgh). Sucrose gradient sedimentation fractionates chromatin fibres based on two parameters: their mass (ie DNA length and mass of protein) and their hydrodynamic shape (ie conformation or structure of the fibre). Since we know that chromatin from the two cell lines is likely to have a similar unit mass, if the DNA length of bulk chromatin in each fraction is similar between the two cell lines, it suggests that bulk chromatin from the two cell lines is similarly compacted (ie has the same conformation). By calculating the DNA size corresponding to the intensity peak of each fraction on the gels, I plotted a graph of bulk DNA size against fraction number and confirmed that the profiles of the two cell lines were very similar (*Figure 5.3B*). Taken together, these analyses suggest that although HaCaT and 293 cells appear to exhibit differences in chromatin compaction at the level of interphase organisation (see *Chapter 4*), there are no obvious differences in bulk 30nm fibre conformation between the two cell lines.

5.2 The Overall Secondary Chromatin Structure of the EDC is Similar to its Surroundings

I had hypothesised that, upon activation, the EDC would undergo a general decondensation. My results in *Chapter 4* suggested this might not be the case because cytological compaction across the locus was not uniform and the most compact portions of the EDC in the control cell lines had a similar compaction levels to the flanking regions (*Section 4.4*, p142). However, tertiary and secondary chromatin structure may not be correlated in the region and I wanted to use sucrose gradient sedimentation to analyse the compaction of the EDC at the level of secondary chromatin structure.

One way of assaying this would be to perform Southern blots on DNA isolated from sucrose gradient fractions and compare the sedimentation of the EDC to bulk chromatin. Such a methodology has been used to demonstrate that the active chicken *β-Globin* locus has a more open conformation than bulk chromatin and that mouse minor satellite is more compact (Gilbert and Allan, 2001;Kimura et al., 1983). However, the size of the EDC makes designing a suitable probe difficult, so instead I elected to perform the analysis using the EDC oligonucleotide microarrays. Rather than comparing the compaction of the EDC to bulk chromatin, I would be comparing its properties to those of the flanking genomic regions.

Each fraction isolated from the gradients contained a mixture of DNA sizes (*Figure 5.3A*). The bulk of chromatin fibres in a particular fraction are of an average length giving the intensity peak measured in the previous section. However, co-sedimenting with these are chromatin fibres of a shorter DNA length which have moved through the gradient faster than expected, implying they have a more compact, streamlined conformation. Conversely there are also DNA fragments of a greater size which are sedimenting higher up the gradient than expected from their mass alone. Therefore, their frictional coefficient must be higher than bulk chromatin, implying a more disorganised, 'open' structure. Previously it has been suggested that the open conformation of a chromatin fragment from the active *β-globin* locus was caused by the presence DNase hypersensitive sites which create discontinuities in the chromatin fibre (Caplan et al., 1987).

By running a single fraction on a pulse field gel, DNA enriched in either open or compact fibres can be extracted from gel slices. Previously, our laboratory performed a genome wide analysis of the distribution of open and compact chromatin from human lymphoblastoid cells using a combination of FISH and microarrays (Gilbert et al., 2004). I proposed to use a similar method to examine the distribution of open chromatin fibres at the EDC (*Figure 5.4*). Therefore, I ran single fractions from the bottom of the gradient on pulse field gels and cut out bands corresponding to DNA twice the size of bulk chromatin. Fractions containing a large average DNA size were chosen because features of secondary structure are

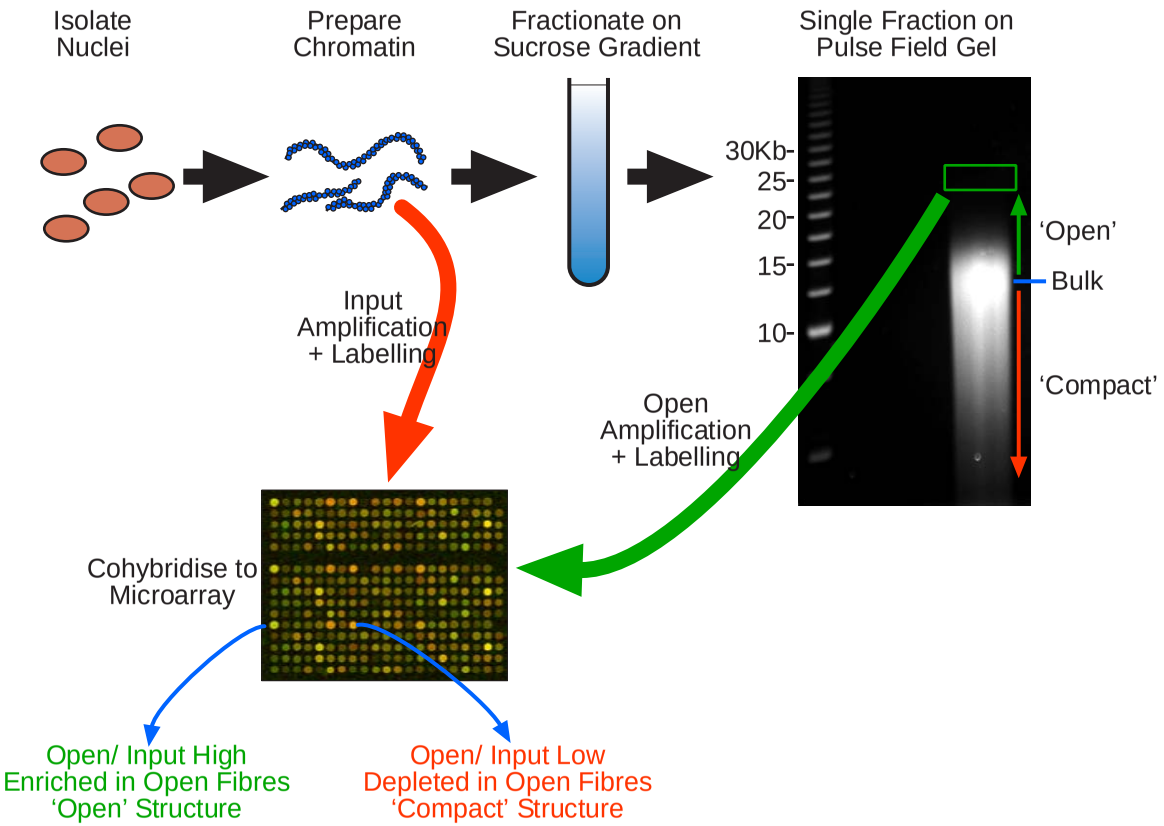


Figure 5.4 Assessing the Distribution of Open Chromatin Fibres Across the Genome

Chromatin is prepared from nuclei by micrococcal nucleas digestion and fractionated on a sucrose gradient. Within a single fraction the bulk of chromatin is of an average size. However, DNA with a greater length is also present and must have been packaged in a more open conformation along with DNA of a shorter length that had been packaged in a more compact way. 'Open' chromatin can be extracted from the gel and hybridised to genomic microarrays together with input material to assay the 'openess' of a particular region.

more likely to be preserved in large chromatin fragments than in small ones. Very low levels of DNA were extracted from these gel slices and therefore I had to amplify it in order to hybridise it to a microarray. I used a commercial Whole Genome Amplification kit (Sigma) which uses chemical shearing and a PCR like method to amplify the DNA. I had previously determined that this method amplified complex mixes of DNA in a representative manner (*Section 2.3.8,* p59).

After amplification and labelling, open chromatin was hybridised to EDC oligonucleotide arrays, using input chromatin to provide a normalisation control and to correct for the differential abundance of DNA sequences in the chromatin loaded on the gradient. Therefore, the open/input ratio of a particular array probe was a measure of the degree of enrichment in open chromatin fibres of a particular sequence, in other words the 'openness' of a particular region of the genome (*Figure 5.4*). It is important to note that from this analysis I cannot infer the actual conformation of the chromatin fibre in a particular genomic region, only that it is more likely to have an open conformation than a region which has a lower open/input ratio.

For each amplified fraction of open chromatin, I labelled and hybridised it twice to the arrays in opposite dye orientations before excluding probes from the analysis based on the dye swap (see *Section 2.12.1,* p100) and calculated a mean $\log_2$ open/input ratio from the two hybridisations. Then, I generated a further amplified open chromatin fraction from an independent sucrose gradient and hybridised it to the arrays in the same fashion as above. For both cell lines, the mean $\log_2$ ratios between the two biological replicates were correlated (HaCaT Pearson's R^2=0.523 and 293 R^2=0.549). This correlation was lower than observed between two replicates of a technique such as ChIP (see *Section 5.9* below) but higher than that observed between replicates of control self hybridisations (R^2=-0.086, this probably represents the level of correlation expected from random error). Therefore, it suggests that a consistent pattern of secondary chromatin structure must exist in each cell line but, perhaps due to the comparatively greater number of steps involved in the experiment where variation can be induced, the results are more variable and we observe a somewhat lower correlation between experimental replicates than when analysing chromatin structure through other means.

Next I compared secondary chromatin structure in the EDC to the surrounding genomic regions. On the EDC oligonucleotide microarrays, probes are tiled across the EDC and ~700Kb of its immediate surroundings (Flanking 1 and 2, *Figure 5.5A*). Therefore the secondary chromatin structure of the EDC could be compared to the flanking regions by comparing the distribution of $\log_2$ open/input ratios at each of three regions. If the EDC was more compact than its surroundings, we would expect that less open chromatin fibres corresponding to the region would have been purified from the gradient and therefore, on average log ratios from the EDC would be lower than from the flanking regions causing the distribution to shift towards negative values. Conversely if the EDC is more open than its surroundings, more open fibres will be purified corresponding to the region resulting in higher log values and a shift in the distribution towards positive values.

Boxplots of the distributions of mean $\log_2$ open/input from all four hybridisations showed that the EDC was not, as a whole, more compacted or decondensed relative to the flanking regions in either cell line (*Figure 5.5B*). However, the distribution of values from flanking region 1 was found to be significantly different from that of the EDC in HaCaT cells (KS-test, p=0.043). This appeared to be because the distribution of values observed for the EDC was wider than that of the flanking region. This suggests that the overall secondary structure of the EDC in HaCaT cells was not more open or compact than in flanking region 1, but is perhaps more variable. The distribution of each region in 293 cells was significantly wider than that in HaCaT cells (KS-tests). However, due to the number of variables involved in the experiment, it is difficult to draw any conclusions from this (for example, it could be caused by some difference in the efficiency of amplification or the hybridisation). Therefore, overall the EDC does not appear to be more compact or open than the surrounding genomic regions in HaCaT or 293 cells. This is agrees with my analysis of interphase compaction in *Chapter 4*, which suggests that the tertiary structure of the active EDC is also not more generally open than its surroundings.

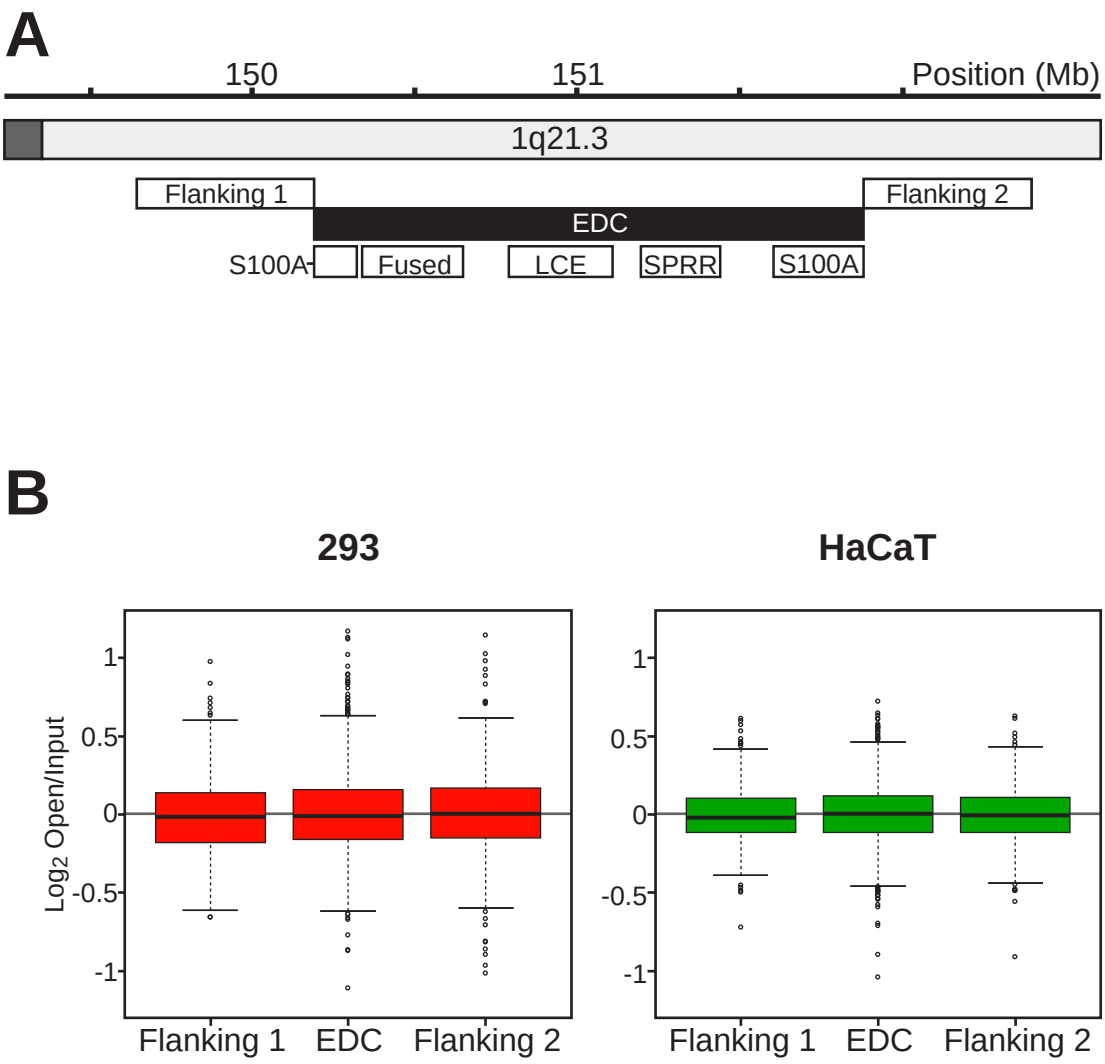


Figure 5.5 The Chromatin Structure of the EDC does not differ from its Surroundings

A. Schematic of the region around the EDC showing the regions tested. The EDC is shown along with its gene clusters. The two flanking regions lying outside of the EDC but still represented by a high density of probes on the EDC oligonucleotide arrays. All positions taken from NCBI build 36.

B. Comparison of the chromatin structure of the EDC and flanking regions in 293 and HaCaT cells. For each boxplot the median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles).

5.3 Does Chromatin Fibre Structure Correlate with Interphase Organisation?

Using a similar sedimentation method and assaying chromatin fibre structure across the human genome using a low resolution BAC tiling array, it has been shown that a region enriched in open chromatin fibres has a lower interphase compaction (as assayed by FISH) than one that is depleted in open chromatin fibres (Gilbert et al., 2004). I have measured both secondary chromatin structure and interphase chromatin compaction (see *Chapter 4*) to higher resolution at the EDC (on the BAC arrays probes were spaced ~1 every Mb). Therefore, I can determine whether a relatively simple relationship between secondary and tertiary chromatin organisation also exists at the EDC. To compare the two, I passed a 500Kb sliding window across the region and calculated mean $\log_2$ open/input ratios for each interval. By plotting the resulting profiles for both cell lines, I found similarities between the secondary structure profile (*Figure 5.6A*) and interphase compaction measured by FISH in 500Kb intervals (*Figure 4.8*, p147). For example, the region covered by probe pair 4 (which covers the distal *S100A* cluster) was the most cytologically decondensed in both cell types and the greatest enrichment of open fibres across a 500Kb interval also occurred within this region.

To test the correlation more comprehensively I calculated the mean $\log_2$ open/input ratio for each of the regions tested by interphase FISH and plotted it against the mean d^2/Mb for each region (*Figure 5.6B*). In both 293 and HaCaT a linear relationship appeared to exist between the two. However, discrepancies can be found, for example the interphase compaction of probe pair 5, which lies just outside the EDC, was similar to that of probe pair 4 in HaCaT cells, but it was much less enriched in open chromatin fibres. Because of this and the relatively low number of datapoints the correlation in both cell lines was rather low (*Figure 5.6B*). Overall however, from this result and a previous study from our laboratory (Gilbert et al., 2004), I conclude that a fairly simple, direct correlation exists between a region's secondary chromatin structure and its interphase compaction.

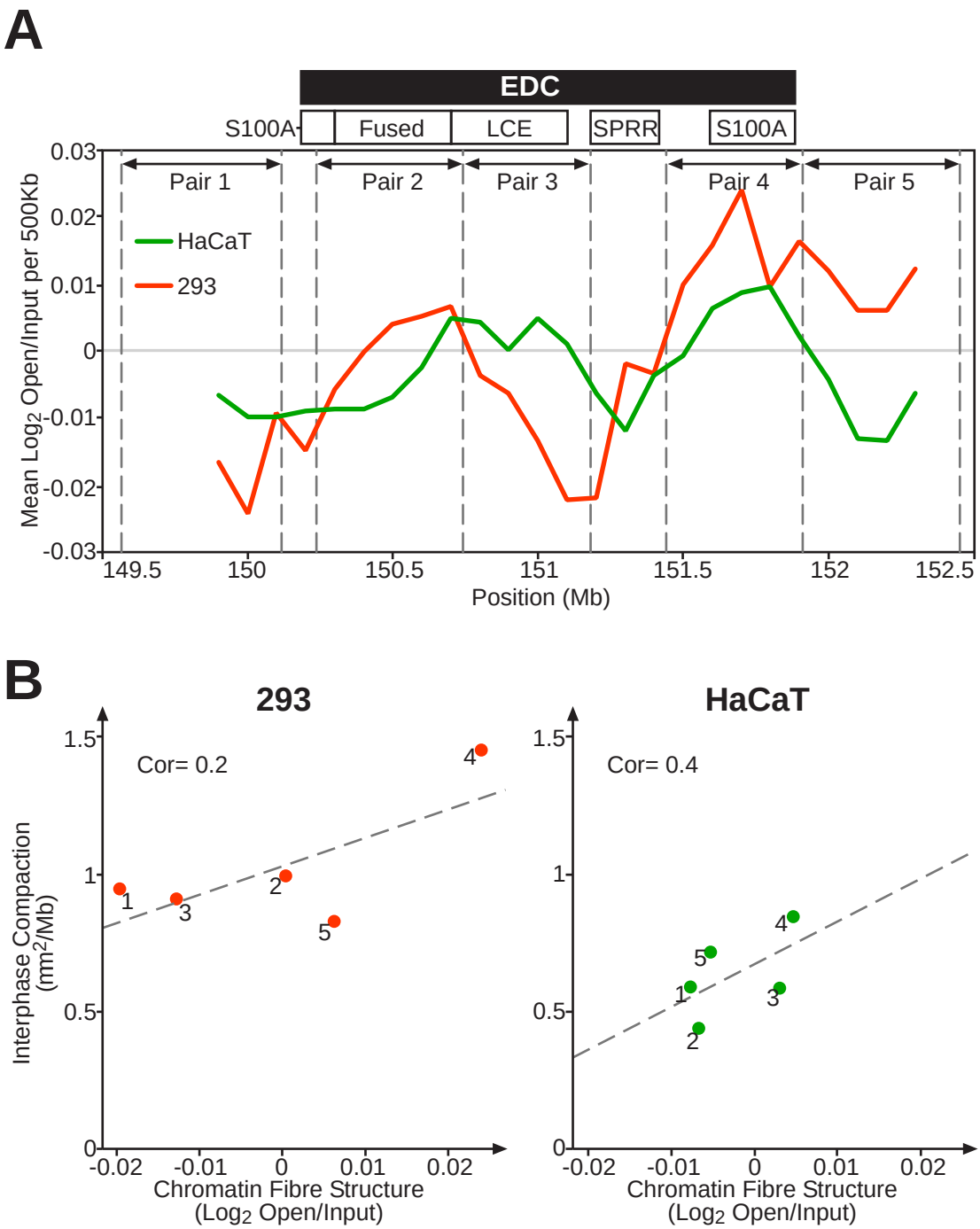


Figure 5.6 The Relationship Between Interphase Compaction and Chromatin Fibre Structure at the EDC

A. Plot of Chromatin Fibre Structure against genomic position across the EDC. A 500Kb sliding window was passed along the EDC in 100Kb increments to calculate the mean $\log_2$ open/input ratio. The position of the EDC and its gene clusters are indicated along with the regions whose interphase compaction was assayed by FISH in *Chapter 4* (see *Section 4.4*). Genomic positions are all taken from NCBI build 36.

B. Correlation between chromatin fibre structure and interphase compaction. The interphase compaction values measured in *Chapter 4* were plotted against the mean $\log_2$ open/input value for the probes within each region. A line of best fit was drawn for each cell line and the Kendalls correlation calculated (see *Section 2.10.2*, p 95). Beside each data point the number of the region assayed by FISH is given.

5.4 Secondary Chromatin Structure Correlates to Gene Density Within the Active EDC

Our previous genome wide analysis of secondary chromatin structure in lymphoblastoid cells showed that, at the scale of genome bands (ie >1Mb), gene dense regions were generally more enriched in open chromatin fibres than gene poor regions (Gilbert et al., 2004). I wondered whether this was also true at the lower genomic scales examined here, so I calculated the gene density across the EDC using a 250Kb sliding window. Similarly, I calculated mean $\log_2$ open/input values using a sliding window of the same size. The secondary chromatin structure profiles of the two cell lines at this scale (*Figure 5.7A*) were similar to their profiles using a 500Kb sliding window (*Figure 5.6A*). When I compared the gene density profile to the secondary structure profile of the two cell lines, I found similarities in the HaCaT profile to that of gene density (*Figure 5.7B*), for example, the most distal part of the region (at ~152.5Mb) was relatively decondensed in both cell lines and also gene rich. The most gene dense part of the region (the distal *S100A* clusters) was enriched in open chromatin structure in HaCaT cells.

Across the whole region, the 250Kb sliding window profiles of chromatin structure in HaCaT cells and gene density were significantly correlated, whereas the secondary chromatin structure profile of 293 cells was not (*Table 5.1*). I wondered whether this might be due to a reorganisation of the secondary structure of the active EDC. Therefore, I calculated the correlation between secondary chromatin structure and gene density solely within the EDC and found again that in HaCaT cells the two significantly correlated but that in 293 cells they did not (*Table 5.1*). In the regions flanking the EDC, there was a positive correlation between gene density and secondary chromatin structure in both cell lines, but again it was only significant in HaCaT cells (*Table 5.1*). The profiles of chromatin structure for HaCaT and 293 cells were not correlated with each other, either across the whole region or within the EDC (Kendall's 0.051 and 0.047 respectively). Therefore, the active and inactive EDC have a different secondary chromatin structure and the organisation of the active EDC reflects gene density whereas that of inactive EDC does not.

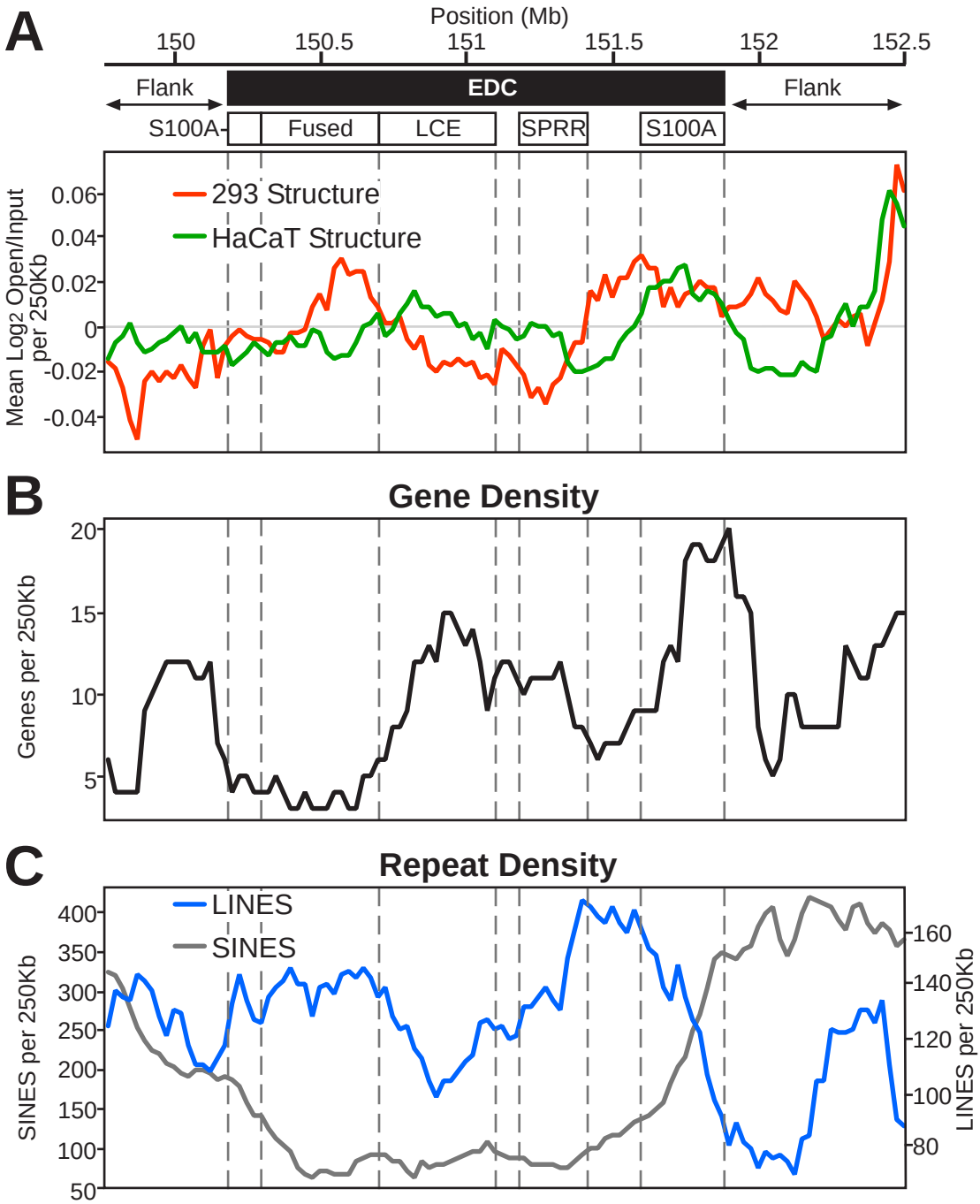


Figure 5.7 Correlation between Secondary Chromatin Structure and Gene Density

A. Plot of chromatin fibre structure against genomic position across the EDC. A 250Kb sliding window was passed along the region in 25Kb increments to calculate the mean $\log_2$ open/input ratio. The position of the EDC and its gene clusters are indicated. Genomic positions are taken from NCBI build 36.

B. Plot of gene density across the EDC. Gene density was calculated using a 250Kb sliding window as for panel A. Gene positions were taken from the midpoints of the genes on the Refseq annotation.

C. Plot of repeat density across the EDC. SINE and LINE density was calculated using a 250Kb sliding window as for panel A. LINES are plotted in Blue and their scale is on the right hand axis, SINES are in Grey and their scale on the left hand axis. Repeats positions were identified from their midpoint on the Repeat Masker annotation.

Table 5.1

Correlation between Gene Density and Secondary Chromatin Structure at a 250Kb Scale around the EDC

Mean gene density and $\log_2$ open/input ratios were calculated in 250Kb intervals using a sliding window across the region (for the plots of the data see *Figure 5.7*). The Kendall's correlation between gene density and mean $\log_2$ open/input ratios was then calculated. The Kendall's correlation was used because it does not assume normality nor a linear relationship. Correlations were also calculated when the EDC or its flanking regions alone were considered. A p-value was assigned to each correlation by random permutation analysis where the values of each probe within the region were replaced by a sample from the whole array and the analysis repeated 1000 times. The p-value of the correlation was then the percentage of the randomly generated correlations greater to or equal to the observed value.

Cell	Region	Kendall's Correlation (Gene density:$\log_2$ open/input)	p-value
293	Whole	-0.021	0.455
	EDC	-0.180	0.254
	Flanks	0.251	0.139
HaCaT	Whole	0.382	0.01
	EDC	0.454	0.02
	Flanks	0.350	0.05

I also examined how the profile of secondary chromatin structure related to the density of repetitive elements within the EDC. In the human genome in general, LINE elements show a negative correlation with gene density and SINE elements (especially Alu elements which make up the majority of this class of repeat) show a positive correlation with gene density (Lander et al., 2001). The profile of secondary chromatin structure at the EDC in 293 cells was similar to the profile of LINE density across the region (*Figure 5.7C*) and there was a moderate correlation between the two, but this was not quite significant (Kendall's = 0.330 and p=0.070). Secondary chromatin structure in HaCaT cells showed a moderately negative correlation to LINE density but, this was also not quite significant (Kendall's = -0.340 and p=0.056). SINE elements were generally depleted from the region of the EDC and chromatin structure in both cell lines was not significantly correlated to

SINE density within the EDC. Outside the EDC, in the flanking regions, 293 cell chromatin structure was negatively correlated with LINE density (Kendall's = -0.512 and p=0.015) but HaCaT cells are not (Kendall's = 0.337 and p=0.097). Therefore, this suggests that the secondary chromatin organisation of the active EDC in HaCaT cells reflects gene density and secondary structure at the inactive EDC in 293 cells has a different organisation that is in some way related to LINE density. This analysis also suggested that within the EDC, the normal relationship between SINE density and gene density in the genome breaks down (Kendall's =0.165, LINE density and gene density are still negatively correlated, Kendall's =-0.359).

5.5 GC Content Does Not Correlate to Secondary Chromatin Structure at the EDC

We know from the sequencing of the human genome that in general genes are concentrated in regions of the genome that are GC-rich and that LINE elements are in contrast found in GC-poor regions (Lander et al., 2001). Therefore, the fact that in HaCaT cells, secondary chromatin structure is correlated with gene density would suggest that the structure may be related to the GC-content of the region. Likewise the structure of the inactive EDC in 293 cells correlates to some extent with LINE density and may also correlate with GC-content, albeit in the opposite manner to that of the active EDC.

To test this, I calculated GC-content across the EDC using a 250Kb sliding window and the 5bp GC-content annotation from the UCSC browser (March 2006 assembly). GC-content in the region varied from ~40-50% and was weakly correlated to the gene density (Kendall's=0.262), probably because of a dissociation of the two profiles in the region around the LCE and SPRR gene clusters (*Figure 5.8*). I did not find a significant correlation between GC-content and secondary chromatin structure in either cell line. The correlation between GC-content and secondary chromatin structure at the EDC in HaCaT cells was lower than between

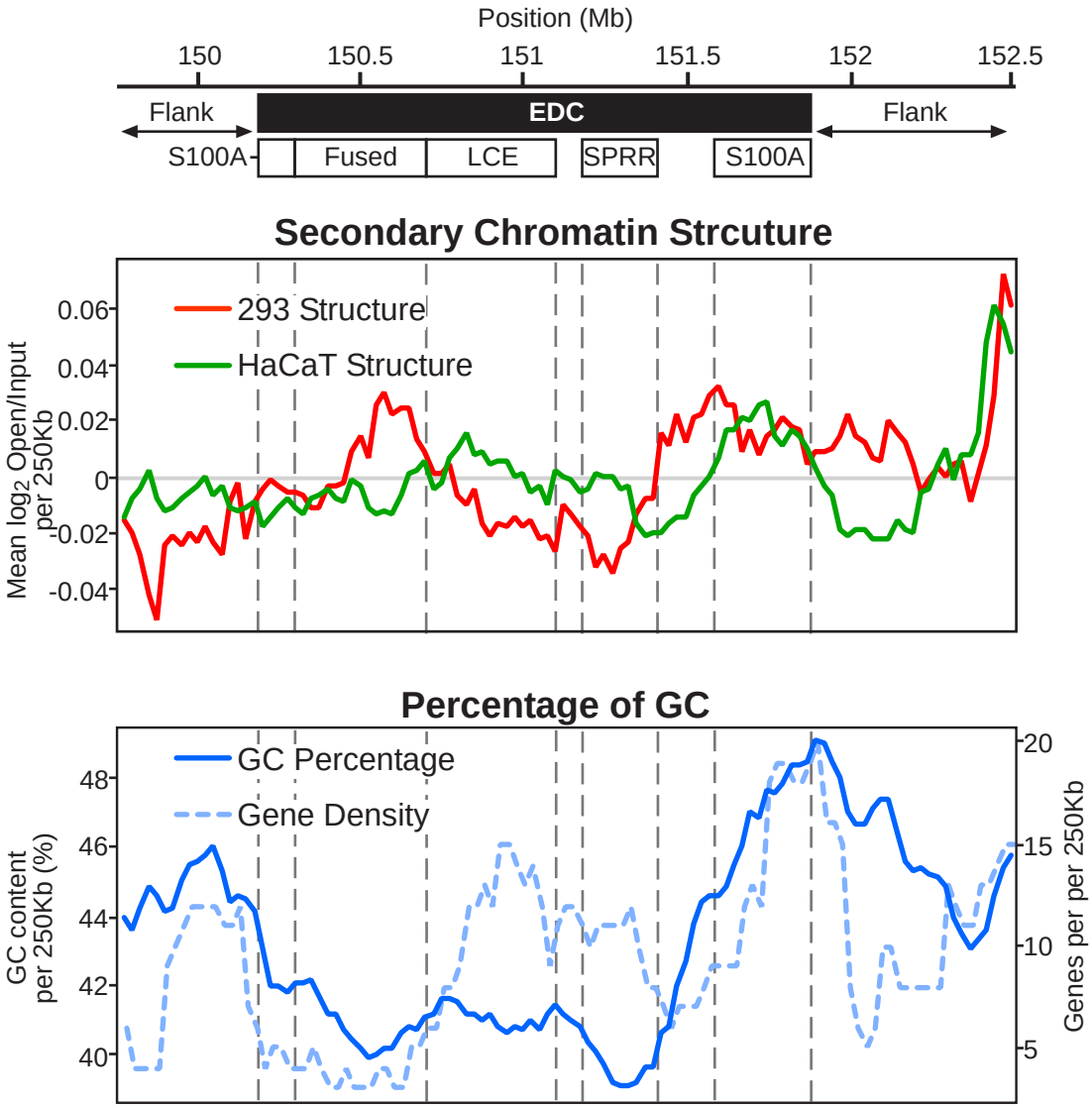


Figure 5.8 Correlation between Secondary Chromatin Structure and GC Content

In the top plot, chromatin fibre structure is plotted against genome position for 293 and HaCaT cells. A 250Kb sliding window was passed along the region in 25Kb increments to calculate the mean $\log_2$ open/input ratio. In the lower plot, the 250Kb sliding window was used to calculate the GC content of the region using the GC percentage annotation from the UCSC genome browser. The gene density calculated by the sliding window is also shown using the Refseq annotation. Above the plots the genomic structure of the EDC and its gene clusters are shown. All genomic positions are from NCBI build 36.

gene density and chromatin structure (*Figure 5.8*, Kendall's=0.296, p=0.082). Surprisingly, GC-content and chromatin structure are positively correlated in 293 cells within the EDC but not significantly (Kendall's=0.309, p=0.104). Therefore, the secondary chromatin organisation of both the active and inactive EDC do not correlate strongly to the GC-content of the region. The organisation of the active EDC is more strongly related to differences in gene density and the inactive EDC correlates weakly with LINE density.

5.6 Genic Regions are Not Enriched in Open Chromatin Fibres at the Active EDC

The correlation between secondary chromatin structure and gene density might be explained if genic regions adopted an open chromatin structure in the active EDC. This may also in part explain the association between gene-rich regions and secondary chromatin structure across the whole genome when examined at large scales (Gilbert et al., 2004). Firstly, I examined the chromatin structure data from both cell lines in the region around the *SPRR* cluster and distal *S100A* gene cluster because this represented a region of transition from a gene-poor to a gene-rich area (*Figures 5.7* and *5.9A*). Genic regions did not obviously appear to be in open chromatin structure, although one of the *PGLYGRP* genes (Peptidoglycan recognition protein) did seem to be in a region of more open structure in 293 cells.

The distributions of $\log_2$ open/input values for probes in genic regions of the EDC and all other probes were similar in both cell lines (*Figure 5.9B*). However, in 293 cells a small but statistically significant shift towards higher values was seen for the genic probes from the EDC (KS-test, p=0.016). However, this would not explain the pattern of chromatin structure seen in these cells, which is better correlated to the density of LINE elements than genes. I also examined the distribution of probes found in all genic regions on chromosome 1 in both cell lines. No shift was seen in these distributions suggesting that in general genic regions do not adopt a more open chromatin structure than the rest of the genome (data not shown).

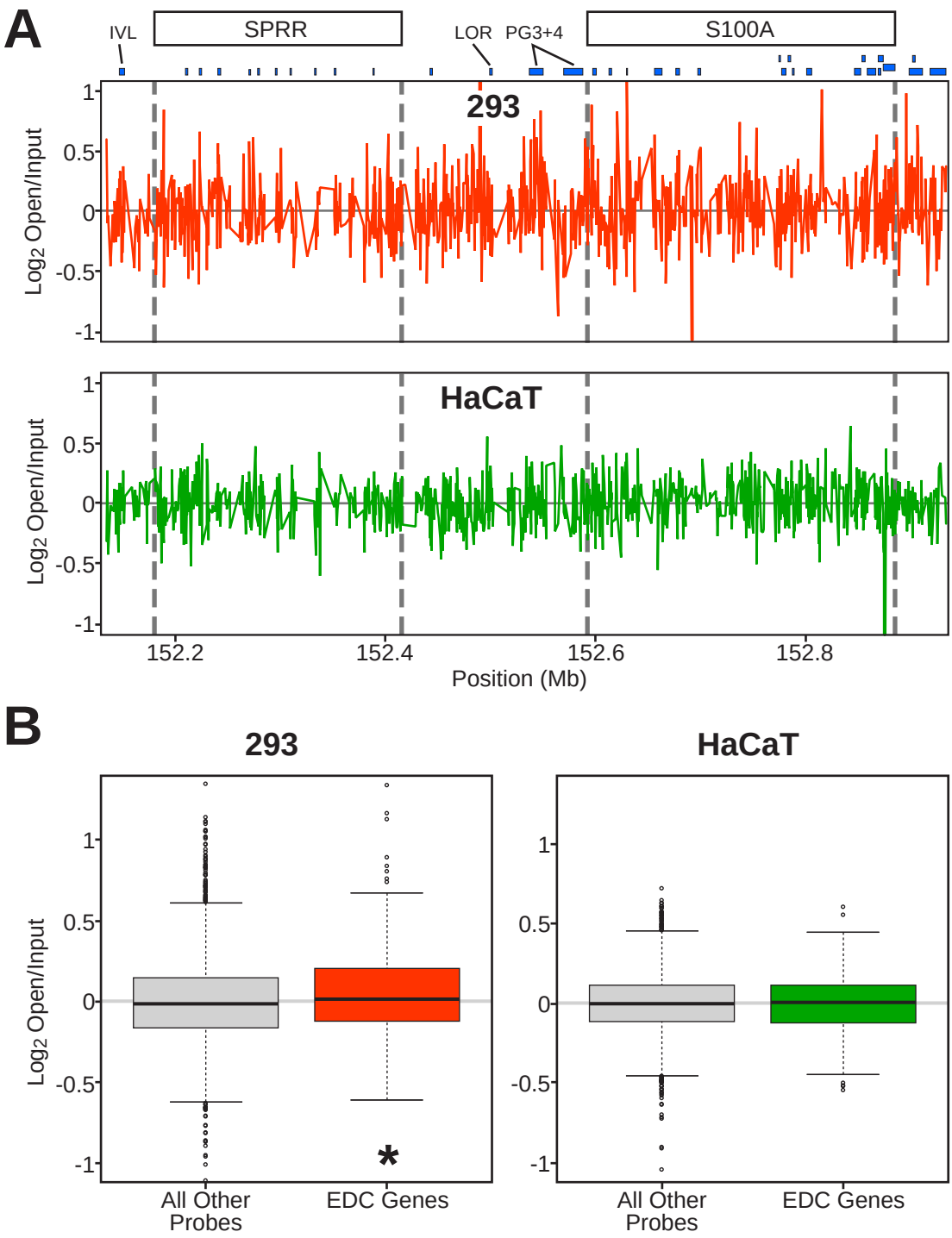


Figure 5.9 Genic Regions Are Not Greatly Enriched in Open Chromatin Fibres

A. Plot of Chromatin Fibre Structure against genomic position in the distal part of the EDC. The positions of the *S100A* and *SPRR* clusters are indicated and below this the positions of annoted genes in the region (Refseq). Positions are all taken from NCBI build 36. The genes *Involucrin* (IVL), *Loricrin* (LOR) and *Peptidoglycan Recognition Protein 3 and 4* (PG3+4) are also indicated.

B. Boxplots comparing the distribution of $\log_2$ open/input values for genes in the EDC and the rest of the genome. Genes are the annotated Refseq genes within the region in NCBI Build 36. Asterisk indicate the distributions is significantly different (KS-Test).

5.7 Weak Sub-100Kb Periodicity of Chromatin Fibre Structure at the EDC

The previous analyses suggested that the active EDC has a different chromatin organisation from that of the inactive locus. By plotting the secondary structure in the region using a 250Kb sliding window, the region appeared to be composed of domains of ~0.5-1Mb in size (*Figure 5.8*). For example, in 293 cells the region around the distal *S100A* cluster could represent a domain of relatively open structure ~1Mb in size and in HaCaT cells there may be a smaller domain, centred on the *S100A* cluster, and flanked by more closed domains. We know that at this scale, secondary chromatin structure reflects higher level chromatin organisation, but how are these chromatin domains built up?

When secondary chromatin structure was examined using a smaller sliding window, the most obvious feature were strong opposing peaks within the fused gene cluster in both cell lines (*Figure 5.10*). However, these corresponded to a region of the genome, proximal to the *filaggrin* gene, which due to the repetitive nature of the underlying sequence, was represented by an extremely low density of array probes (4 in ~50Kb). Therefore, secondary structural data across this region is probably inaccurately measured and should be interpreted with caution.

It could also be seen that within any given region of the locus, chromatin structure was highly variable and perhaps organised into domains of ~50Kb. If these chromatin domains were of a fairly regular size, it may be possible to detect them by analysing the secondary structure for periodicity. In physics, a power spectrum analysis is widely used to examine periodicity. The first step is to transform data using a Discrete Fourier Transformation (DFT) which transforms data taken over discrete intervals, either temporal or spatial, into the frequency domain and isolates individual frequency components of a complex signal for easier analysis. A scaled plot of Fourier transformed data is commonly known as a power spectrum.

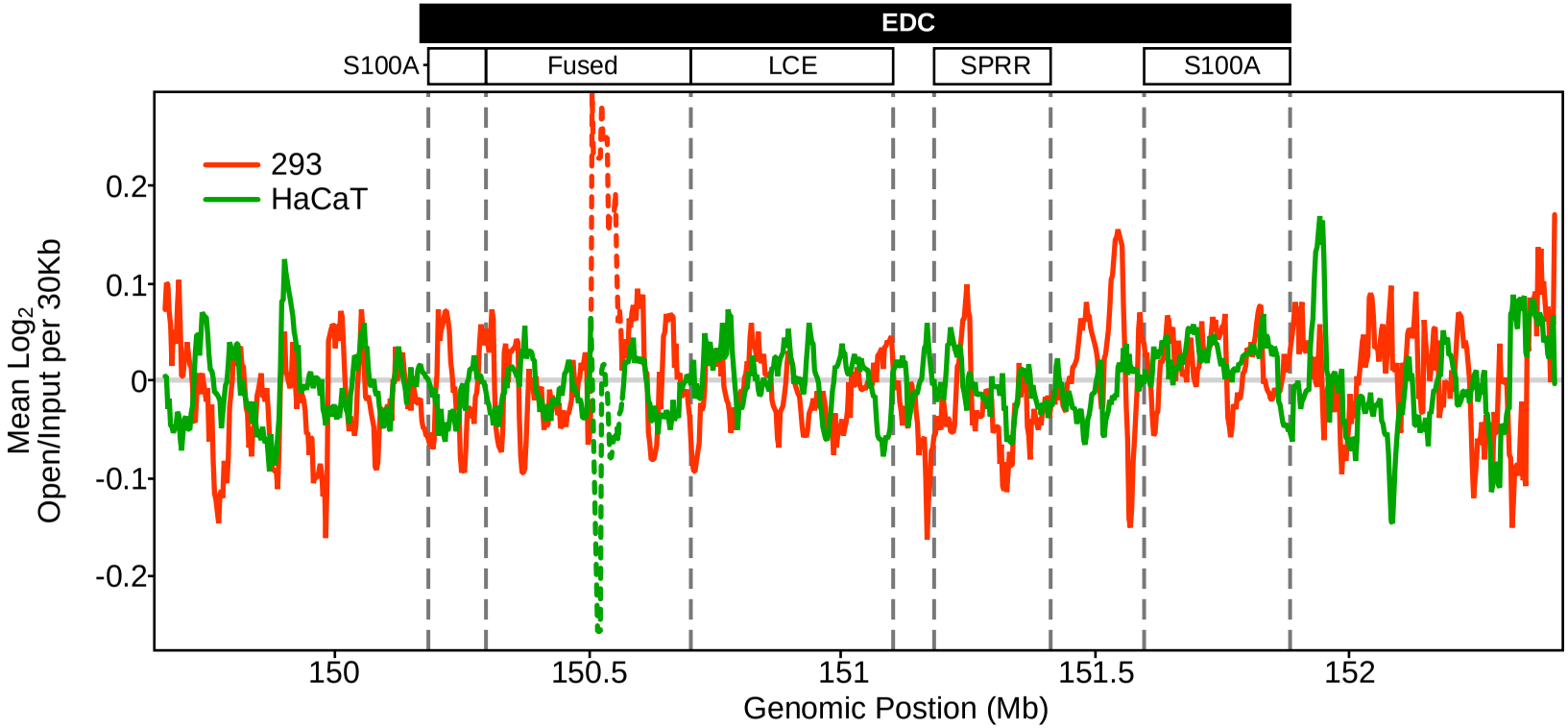


Figure 5.10 Secondary Chromatin Structure at the EDC with a 30Kb Sliding Window

Secondary chromatin structure as measured using sucrose gradient sedimentation and microarrays was analysed using a 30Kb sliding window incremented in 3Kb steps. The EDC and its gene clusters are shown, positions are taken from NCBI build 36. The dashed area of the plot indicates the region around the *filaggrin* gene were the density of oligos is low for around ~50Kb due to the repetiitive nature of the DNA sequence in the region.

The DFT has also been applied to analysis of biological systems. For example, DNA coding sequences exhibit a strong periodicity due to the presence of codons and this has been used in the prediction of genes (Tiwari et al., 1997). In the field of chromatin, autocorrelation analysis and Fourier transformations have been used to examine the periodic placement of nucleosomes and preferential nucleosome placement motifs (Davey et al., 1995;Widom, 1996). It has also been used to detect periodicity in Single Nucleotide Polymorphism density around transcription start sites (Higasa and Hayashi, 2006).

As previously mentioned, this analysis requires data to be sampled at discrete regularly spaced intervals (in this case genomic distance). Due to the constraints of designing probes within the genome, my data is not sampled at a regular interval. To achieve regular data spacing, I used locally weighted scatterplot smoothing (loess) to model the data from individual probes across the region of the EDC in both 293 and HaCaT cells (ie the region of high probe density, EDC + flanks in *Figure 5.5A*, see *Section 2.12.5*, p107). The model was then used to predict evenly spaced datapoints every 1Kb. I adjusted the parameters of the model empirically to attain a good fit to the original data (*Figure 5.11*). In essence, this involved the adjustment of the smoothing level of the model.

I then performed a Fourier transformation on the modelled data from both cell lines, plotting the resulting amplitudes against frequency (*Figure 5.12A*). Similar plots were seen with both cell lines, and there appeared to be peaks in the lower frequency range of the graph (ie $<0.1Kb^{-1}$ or period > 10Kb, arrows *Figure 5.12A*). To examine this further, I replotted the data in the frequency interval 0.01 to $0.1Kb^{-1}$ (period 100Kb and 10Kb respectively) revealing a number of potential amplitude peaks in both cell lines (*Figure 5.12B*).

To assess the biological relevance of the observed peaks I used a random permutation method to apply a significance value to each Fourier transformed datapoint (see *Section 2.12.5*, p107). This process generated a dataset of 1000 control amplitudes for each frequency by random permutation. When the median control amplitude was plotted against frequency, its profile closely followed that of the observed data for both cell lines (*Figure 5.12B*, blue lines).

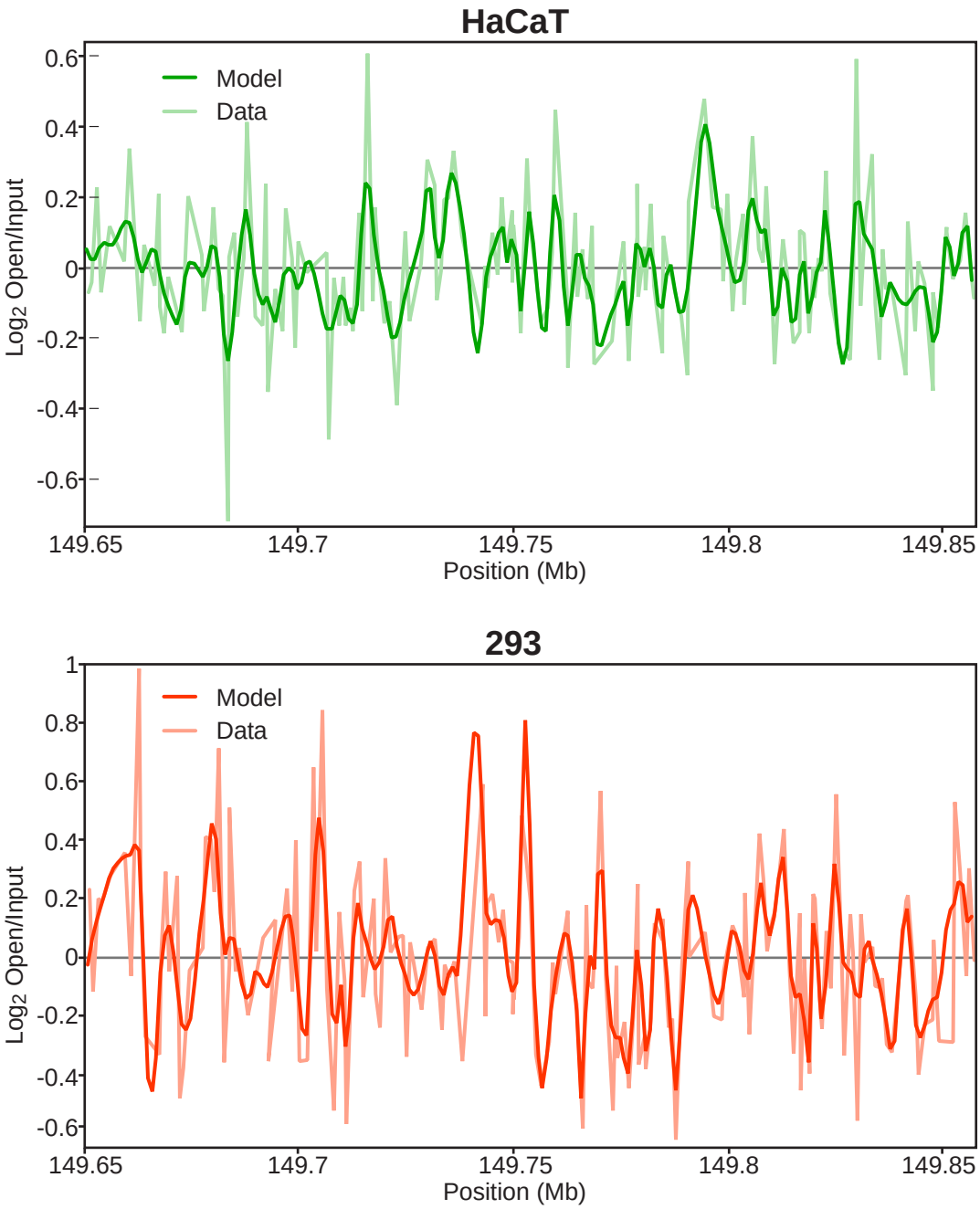


Figure 5.11 Modelling of Sucrose Gradient Data for Periodic Analysis

Data generated from the analysis of open chromatin fibres was modelled using locally weighted scatterplot smoothing. The model was then used to predict data points based on a regular spacing of 1Kb which is required for power spectrum analyses. Scatterplots of log2 Open/Input are shown for a short region of the data (200Kb) to compare the model and the original data for both cell lines. The settings used were determined empirically to give the best representation for the original data. The only factor changed was the percentage of points considered in the modelling of each data point, in the R software, the span, which was set to 0.0025 (in essence, a higher value results in a more highly smoothed plot).

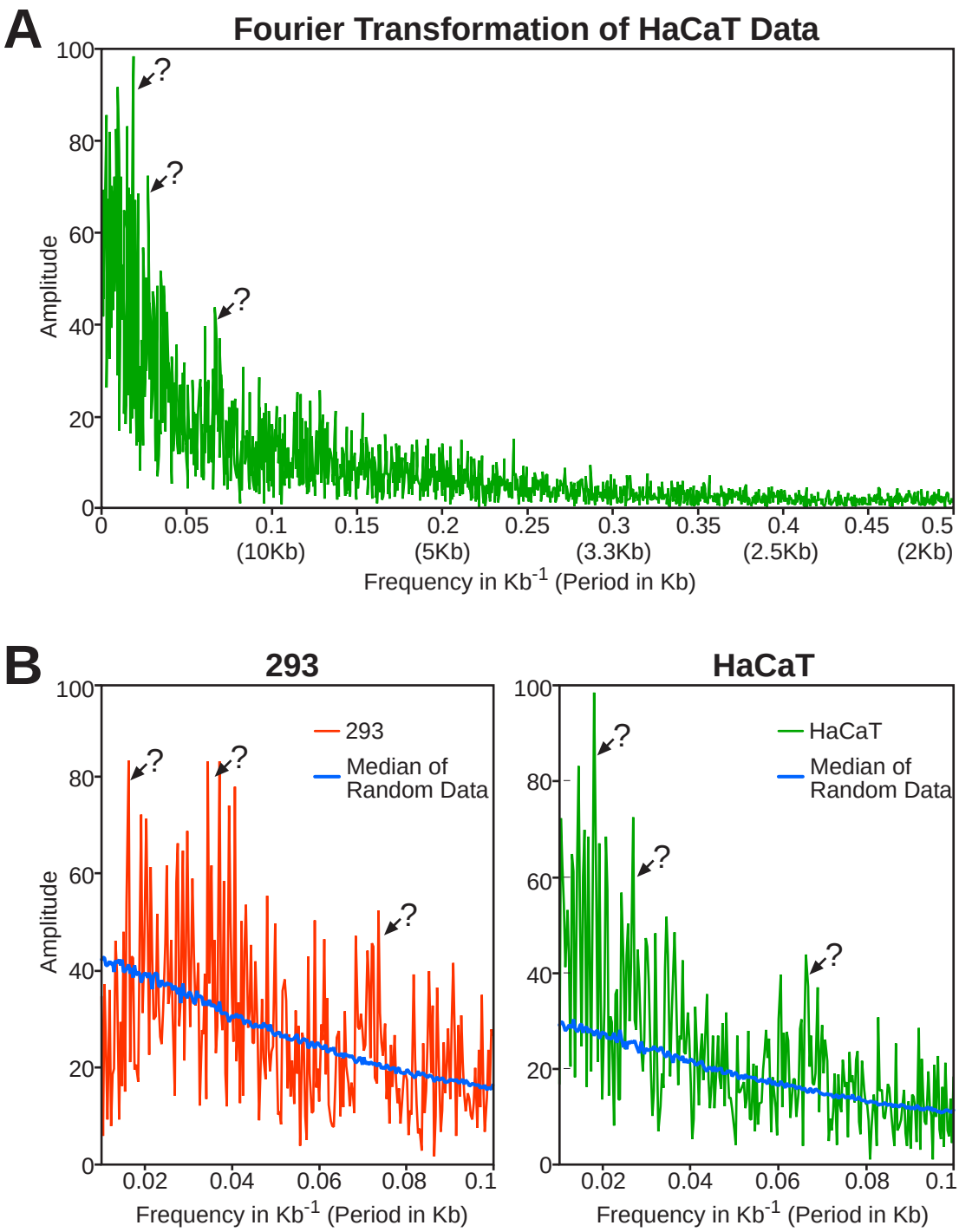


Figure 5.12 Fourier Analysis of 30nm Structure Across the EDC

A. Fourier transformation of modelled Open/Input values from Sucrose Gradient analysis of chromatin structure across the EDC. Microarray data was modelled using a loess function onto a 1Kb interval, it was then Fourier transformed and the variation in frequency plotted against amplitude. The periods or wavelengths corresponding to some frequencies are indicated in brackets (in Kb). The arrowed peaks indicate possible incidences of periodicity in the data.

B. As for A, but for both cell lines and for the frequences 0.01 to 0.1 (ie periods of 100Kb and 10Kb respectively). In addition data from 1000 random samples of Open/Input values from the whole array were also fourier transformed and their median plotted (the blue line in both cases). The same peaks are indicated as in A. for HaCaT cells.

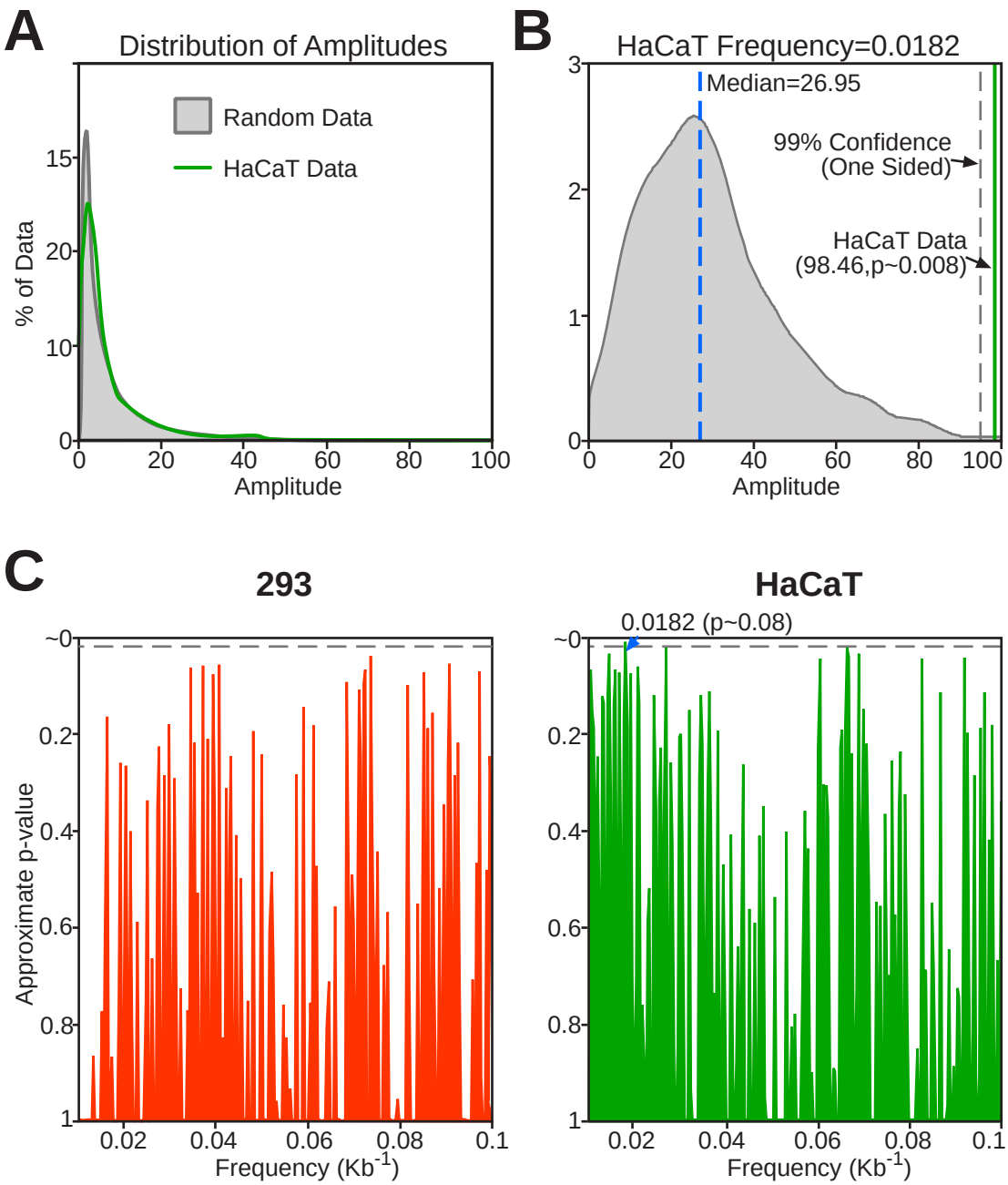


Figure 5.13 Analysis of Periodicity in 30nm Chromatin Structure

A. Histogram showing the distribution of amplitudes seen in 1000 random datasets and in the data from HaCaT cells.

B. Plot of the distribution of amplitudes observed in the random data set for one frequency. The median and one sided 99% percentile (ie 99% percentile of those values >= median) of the distribution are indicated. The 99% percentile represents the 99% confidence interval. From the distribution a p-value can be calculated for amplitudes observed in the HaCaT data, in this case p~0.008.

C. Plot of approximate amplitude p-values by frequency for frequencies between 0.01 and 0.1 (Periods 100Kb and 10Kb respectively). The p-values are calculated from random data sets as shown in B. The 0.01 significance level is indicated by the dashed line and the significant value shown in part B. No other significant values were seen over these frequency intervals.

In addition, the distribution of amplitudes in the observed data for HaCaT cells was not significantly different from the distribution of amplitudes in the entire control dataset (*Figure 5.13A*). Therefore, most peaks were unlikely to represent biologically relevant periodicity in the data. To determine which, if any, peaks were relevant, I employed a simple method to assign p-values to the observed amplitude at each frequency. For each frequency I first determined which control amplitudes that were greater than the control median. The p-value was then the percentage of these amplitudes that were equal to or greater than the observed amplitude at this frequency (*Figure 5.13B*).

Table 5.2

Significant Frequencies and Periods found by Fourier Analysis of Chromatin Structure on Microarrays

Microarray data from the hybridisation of open chromatin fibres purified on sucrose gradients was modelled and Fourier transformed. Random data sets were used to identify amplitude peaks and their corresponding frequencies which were significant above random noise (*Figures 5.12* and *5.13*). Frequency is in Kb^{-1} and Period is in Kb. Separate tables are given for the two cell lines.

293		
Frequency (Kb^{-1})	**Period (Kb)**	**Significance**
0.14139942	7.07	0.001
0.16909621	5.91	0.001
0.31268222	3.20	0.001
0.34766764	2.88	0.001
0.35021866	2.86	0.001
HaCaT		
Frequency (Kb^{-1})	**Period (Kb)**	**Significance**
0.01822157	54.88	0.008
0.12791545	7.82	0.002
0.15269679	6.55	0.01
0.21100583	4.74	0.004
0.24161808	4.14	~0.000
0.35641399	2.81	0.002

Using a 1% significance cutoff, a number of different significant frequencies were observed in both cell lines (*Table 5.2*). Most of these were of a high frequency and therefore low period, indicating periodicity of the data across the range of 1Kb to 10Kb. However, in HaCaT cells a significant peak was observed at a frequency of 0.0182 (*Figure 5.13B* and *C*). This indicates a periodicity in the HaCaT data set of ~54.88Kb but no similarly significant periodicity was found in 293 cells (*Figure 5.13C*). Therefore, there is some evidence for the existence of regularly sized chromatin domains in HaCaT cells of ~50Kb in size. However, the lack of a similar observation in 293 cells suggests this may be a particular feature of HaCaT chromatin in the region of the EDC. This may be explained if irregularly sized domains generally exist in the cell but that at the EDC in HaCaT cells, they are of a more regular size than in 293 cells.

5.8 Correlating Primary and Secondary Chromatin Structure

Histone modifications are the most commonly studied aspect of chromatin structure. They may mediate direct effects on chromatin fibre structure or alternatively, they may instead recruit other proteins to chromatin which then modulate the chromatin structure (the 'histone code' hypothesis (Turner, 1993). There have been a number of experiments examining the effects of histone modifications on chromatin fibre structure *in vitro* (Garcia-Ramirez et al., 1995;Tse et al., 1998), but nobody has yet examined how the pattern of histone modifications in a cell relates to secondary chromatin structure.

Chromatin Immunoprecipitation or ChIP is the most commonly used assay for determining histone modification patterns. It can be performed in two ways, firstly cells are crosslinked, usually with formaldehyde, before the chromatin is fragmented by sonication (Crosslinking ChIP, developed by, Jackson, 1978;Solomon and Varshavsky, 1985;Solomon et al., 1988). Alternatively, chromatin fragments are produced by digestion with an enzyme, usually MNase (Native or NChIP, first developed by, Hebbes et al., 1988). Epitopes of interest are

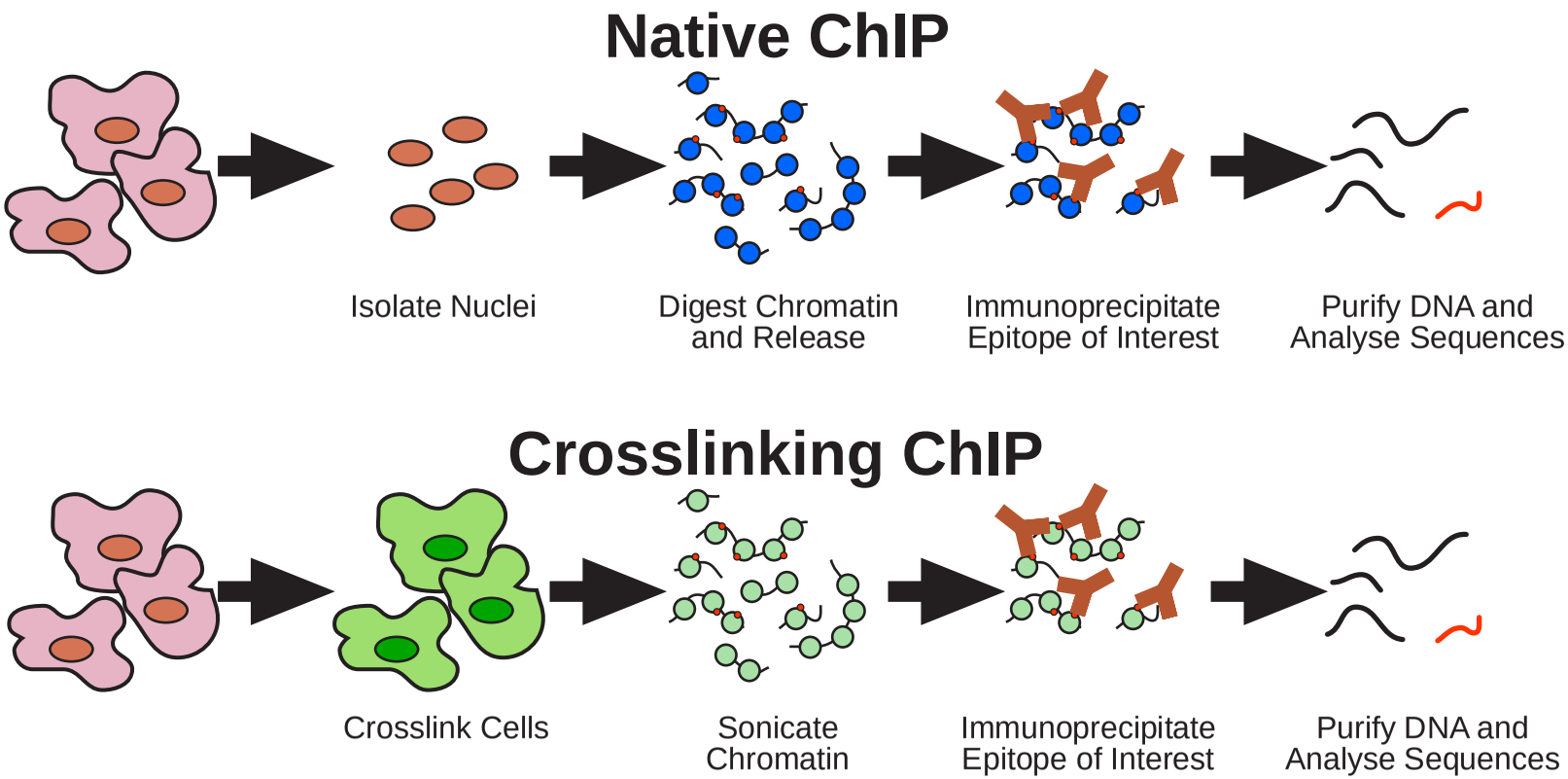


Figure 5.14 Comparison of Native ChIP and Crosslinking ChIP

Comparison of the two methodologies for performing Chromatin Immunoprecipitation (ChIP). The major difference is that in crosslinking ChIP, nuclei are crosslinked (usually with Formaldehyde) and sonicated to prepare chromatin and in Native ChIP the DNA is digested and released gently. In addition to purify DNA from precipitated material, crosslinks have to be first reversed (usually by heat treatment). Assement of Immunoprecipiated DNA is usually by PCR, Southern Blot or Hybdrisiation to Microarrays.

isolated from chromatin fragments produced by both methods and the DNA sequences associated with them are analysed (*Figure 5.14*).

Crosslinking ChIP is more widely used, mainly because it allows the study of any protein, not just those tightly associated with DNA. However, particularly for the study of modifications of core histones, N-ChIP has several advantages. Firstly, a large number of proteins are bound to histones, and many recognise and bind directly to histone modifications themselves. These will remain associated after crosslinking and obscure the modification from antibodies. In addition, formaldehyde crosslinking affects lysine residues and many, but not all, core histone modifications also occur at lysine residues (O'Neill and Turner, 2003;Orlando et al., 1997;Solomon and Varshavsky, 1985). This can alter the structure of the epitopes themselves and may prevent recognition by an antibody. Therefore, the crosslinking process results in lower precipitation efficiency and a less representive precipitation of DNA sequences associated with histone modifications. I decided to use N-ChIP combined with the EDC arrays to examine the distribution of histone modifications at the EDC.

5.9 NChIP can be Used to Assess H3K9Ac Distribution Across the EDC

A variety of methodologies have been used to prepare native chromatin for precipitation (Hebbes et al., 1988;O'Neill and Turner, 2003). I elected to prepare my chromatin in a manner similar to that used in *Section 5.1* because I already knew the digestion properties of my model cell lines with MNase under these conditions. Therefore, I prepared nuclei from both 293 and HaCaT cells and digested them with MNase, but using a greater quantity of MNase than previously. After digestion and allowing chromatin to release into solution, I purified DNAs from both the supernatant and the nuclear pellet and analysed their size distribution (*Figure 5.15A*). The pellet contained insoluble and higher molecular weight DNA whereas the supernatant contained mono-, di-, tri-, tetra- and penta-nucleosomes which is a suitable size for analysis (O'Neill and Turner, 2003).

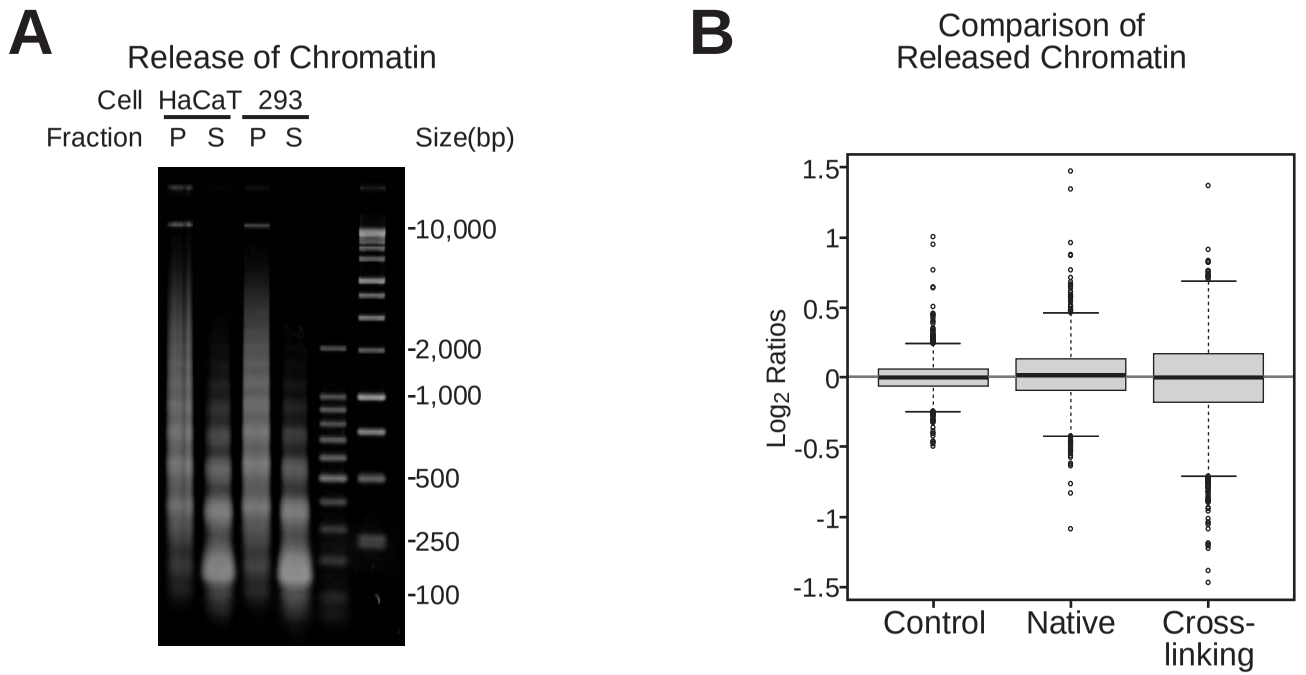


Figure 5.15 Analysis of Soluble Chromatin released by MNase Digestion

A. Ethidium Bromide Stained Gel of DNA isolated MNase digested chromatin from HaCaT and 293 cells. Nuclei were digested with MNase and DNA isolated from both soluble chromatin (S) and the nuclear pellet (P). The sizes of selected DNA markers are indicated.

B. Boxplot of comparing the preparation of Chromatin for ChIP by either MNase digestion (Native) or formaldehyde crosslinking and sonication (Crosslinked). Chromatin was prepared by both methods from HaCaT cells and hybridised against total genomic DNA on EDC oligonucleotide tiling arrays, to test whether the enrichment or depletion of certain sequences had occurred. The control data is from self hybridisations of NChIP Input DNA and is used to estimate the level of experimental error. For each boxplot the median is indicated by the heavy line, the boxed area shows the interval between the 25th and 75th percentiles and outlying observations are indicated by circles (outliers are defined as lying 1.5 times the interquartile distance from the 25th or 75th percentiles).

The preparation of chromatin using a nuclease may result in the depletion of specific sequences which are preferentially digested, or are refractory to digest and remain in the nuclear pellet (O'Neill and Turner, 2003). This may lead to misleading results in the subsequent ChIP, so I compared the population of DNA in soluble chromatin to that of the genome by labelling and hybridising it to the EDC arrays together with genomic DNA. Enrichments and depletions of specific sequences in the input material will therefore be seen as deviations from a 1:1 ratio for the two DNAs (ie a $\log_2$ ratio of 0). I compared its hybridisation pattern to that of a DNA self-self hybridisation (to estimate technical error) and to that of crosslinked and sonicated material with genomic DNA. Boxplots of the mean $\log_2$ distributions from two, normalised and dye swapped hybridisations showed that the distribution of ratios observed with MNase digested native chromatin was significantly wider than that of the control self hybridisation (*Figure 5.15B*, KS-test, $p<1x10^{-15}$). However, the distribution observed for crosslinked chromatin was even broader (KS-test, $p<1x10^{-15}$). Therefore, although not entirely representative of the genome, surprisingly, input chromatin derived by nuclease digestion appears to be more representive than that derived by crosslinking and sonication. This may be due to an incomplete reversal of formaldehyde crosslinks which would cause any DNA still crosslinked to protein to partition into the phenolic phase during phenol-chloroform extractions used to clean up the DNA prior to analysis (Nagy et al., 2003;Solomon and Varshavsky, 1985).

Anti-H3K9Ac was the best characterised antibody for ChIP in our laboratory (C. Kress, personal communication and Chambeyron and Bickmore, 2004)) and so I decided to use it to test the efficiency of performing ChIP on native chromatin. Therefore, using a protocol previously developed by C. Kress, I performed immunoprecipitations of MNase digested chromatin prepared from 293 cells. After precipitation, washing and elution of protein DNA complexes, I used TCA to precipitate the eluted proteins and analysed them by western blotting using anti-H3K9Ac (*Figure 5.16A*). The H3K9Ac epitope was present in protein immunoprecipitated by the antibody but not protein that was mock precipitated using total rabbit serum.

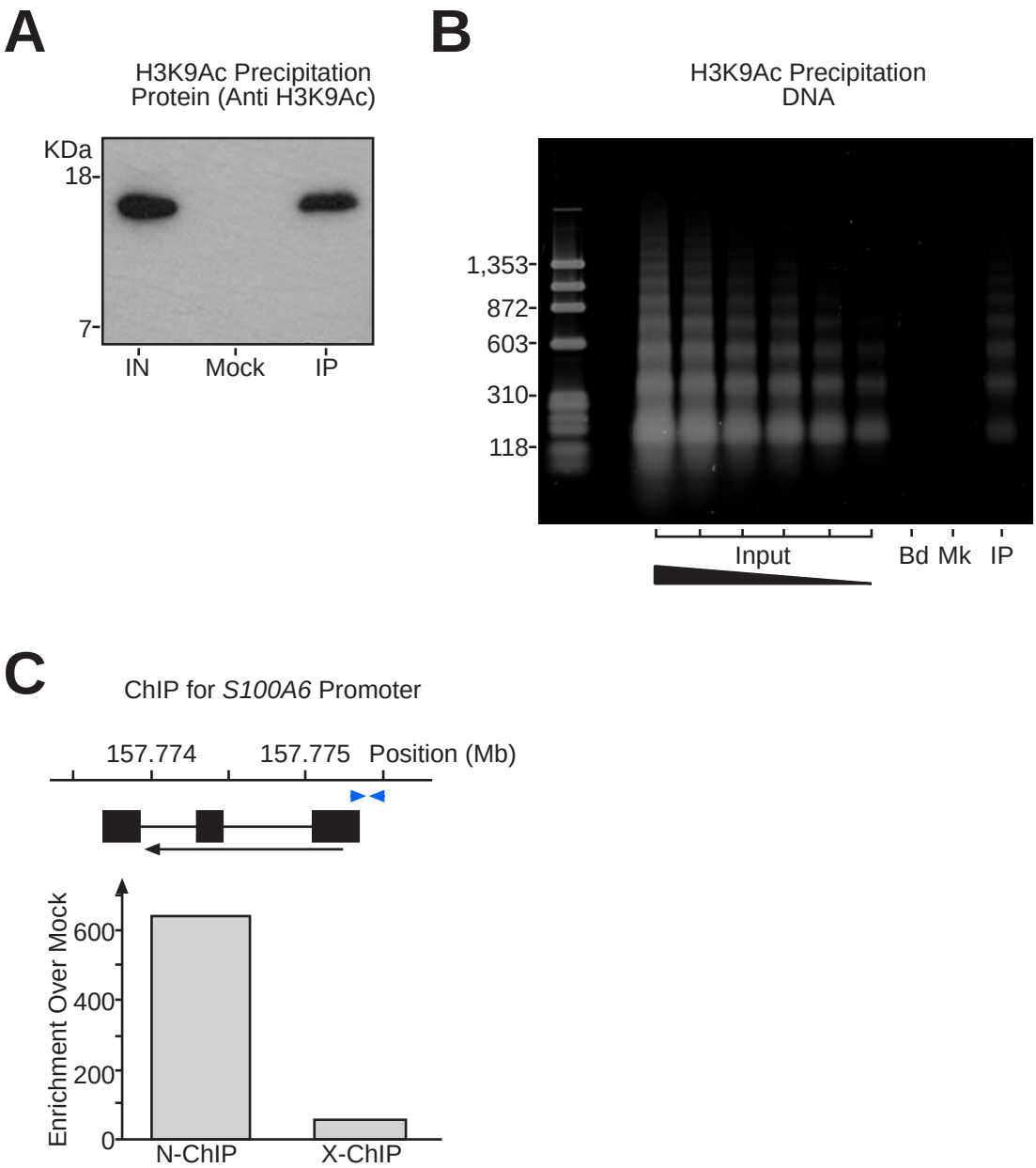


Figure 5.16 Native Chromatin can be Immunoprecipitated

A. Immunoprecipitation of H3K9Ac. Chromatin was prepared from 293 cells and either mock immunoprecipitated using rabbit serum or precipitated with anti-H3K9Ac before being analysed by western blot. using the same antibody. IN = 2% Input, IP = 10% of H3K9Ac precipitated material.

B. Chromatin Immunoprecipitation of H3K9Ac from 293 Cells. Chromatin was immunoprecipitated and purified DNA stained with SYBR Gold. Shown are serial dilutions of input material, precipitation without antibody (Bd), mock precipitation with rabbit serum (Mk) and precipitation with anti-H3K9Ac (IP).

C. Enrichment in Native ChIP (NChIP) vs crosslinking ChIP (XChIP). Real Time PCR was performed on H3K9Ac and Mock precipitated material using primers specific for the *S100A6* promoter. Enrichment over mock values were calculated and plotted. Above the graph the area around the *S100A6* gene is depicted (genomic positions from NCBI build 36). The primers for the PCR are shown as small blue arrows and the direction of transcription indicated below the gene.

The fact that in N-ChIP proteins are not crosslinked to the DNA means that it is possible that histone-DNA interactions had been disrupted during the washing process. I confirmed that this had not occurred by repeating the precipitation and preparing DNA from the eluted material. This DNA was resolved on a gel together with serial dilutions of an aliquot of input chromatin and DNA from control precipitations using no antibody or rabbit serum. I expected very little DNA to be present, so I stained the gel with SYBR Gold which can detect small amounts of DNA (down to ~50pg of Salmon Sperm DNA, data not shown). A nucleosome ladder was seen in the input and H3K9Ac precipitated material but not the control precipitations (*Figure 5.16B*). Immunoprecipitated DNA appeared relatively depleted in mono-nucleosomes, something that has previously been noted by other groups using N-ChIP (O'Neill and Turner, 2003 and references therein). This may in part be explained by the greater probability that a longer fragment will contain an epitope of interest. In addition, perhaps longer fragments containing two H3K9Ac marks are more tightly bound by the antibodies which also have two antigen recognition sites.

Recently, the presence of H3K9 acetylation has been shown at the promoter of the active *S100A6* gene in keratinocytes (Lesniak et al., 2007). To confirm my ChIP was specific I performed Real-time PCR using primers for this sequence. *S100A6* promoter sequences were ~650 times enriched over mock precipitations in an N-ChIP sample from HaCaT cells (*Figure 5.16C*). They were also enriched in DNA immunoprecipitated from crosslinked chromatin but to a far lesser degree. Therefore, although lacking a negative control, my PCR suggests that, as expected, performing ChIP for histone modifications using native chromatin is more efficient than performing ChIP on crosslinked and sonicated chromatin.

To assess the distribution of modifications across the EDC I hybridised H3K9Ac enriched DNA to EDC oligonucleotide microarrays. ChIP-chip is usually, but not always (see (Koch et al., 2007), performed by first amplifying the precipitated material before labelling and hybridisation to arrays (Kim and Ren, 2006). However, amplification can alter the ratios of DNA sequences in a sample and produce artefacts in the final results. I estimated that I had precipitated sufficient quantities of DNA to label and hybridise it to microarrays without prior amplification

(data not shown). Therefore, using random prime labelling I hybridised input material together with anti-H3K9Ac immunoprecipitated DNA (*Figure 5.17A*). I also hybridised 3 control samples: water, DNA precipitated in the absence of antibody and mock precipitated material (all samples same as *Figure 5.16C*). All arrays were hybridised in the same dye orientation (ie Cy3 Input, Cy5 sample) and scanned at the same laser power and photomultiplier gain to allow direct comparisons. A strong signal was produced on the arrays by anti-H3K9Ac precipitated material but not by the controls (*Figure 5.17A*). Quantitation of the average $\log_2$ sample/input ratio for each probe in the hybridisations confirmed that nearly all ratios on the H3K9Ac hybridised arrays were higher than those observed in the controls (*Figure 5.17B*).

In general, two colour ChIP on chip experiments are performed by using a reference sample experimental design (for an explanation of the reference design see**,** *Section 3.2*, p115 or (Churchill, 2002). Therefore, all experimental samples are compared through a common reference sample, usually sheared genomic DNA or input chromatin (Buck and Lieb, 2004). It is preferable to use DNA isolated from input chromatin as this means that the calculated ratios from hybridisations are corrected for any variations in sequence abundance in input chromatin. Direct comparisons between mock and specific precipitations are not usually made because the low DNA levels in mock precipitations results in faint signals and inaccurately measured ratios (Buck and Lieb, 2004). It is generally assumed that the input DNA represents unenriched DNA and that mock precipitated material is similarly unenriched, representing the non-specific precipitation of DNA probably through binding to the beads used in the immunoprecipitation (Kim and Ren, 2006). The fact that more DNA appeared to be precipitated by rabbit serum than beads alone suggested this was not entirely the case (compare beads only and mock in *Figure 5.17B*). To test the assumption I therefore calculated the Pearson's correlation between $\log_2$ H3K9Ac/Input ratios and $\log_2$ H3K9Ac/Mock ratios (calculated from H3K9Ac/Input divided by H3K9Ac/Mock). The correlation between the two was equivalent to that expected between two separate hybridisations of the same samples, which suggests that H3K9Ac/Input ratios are a satisfactory measure of H3K9Ac/Mock ratios (Pearson's, R^2= 0.776, Churchill, 2002).

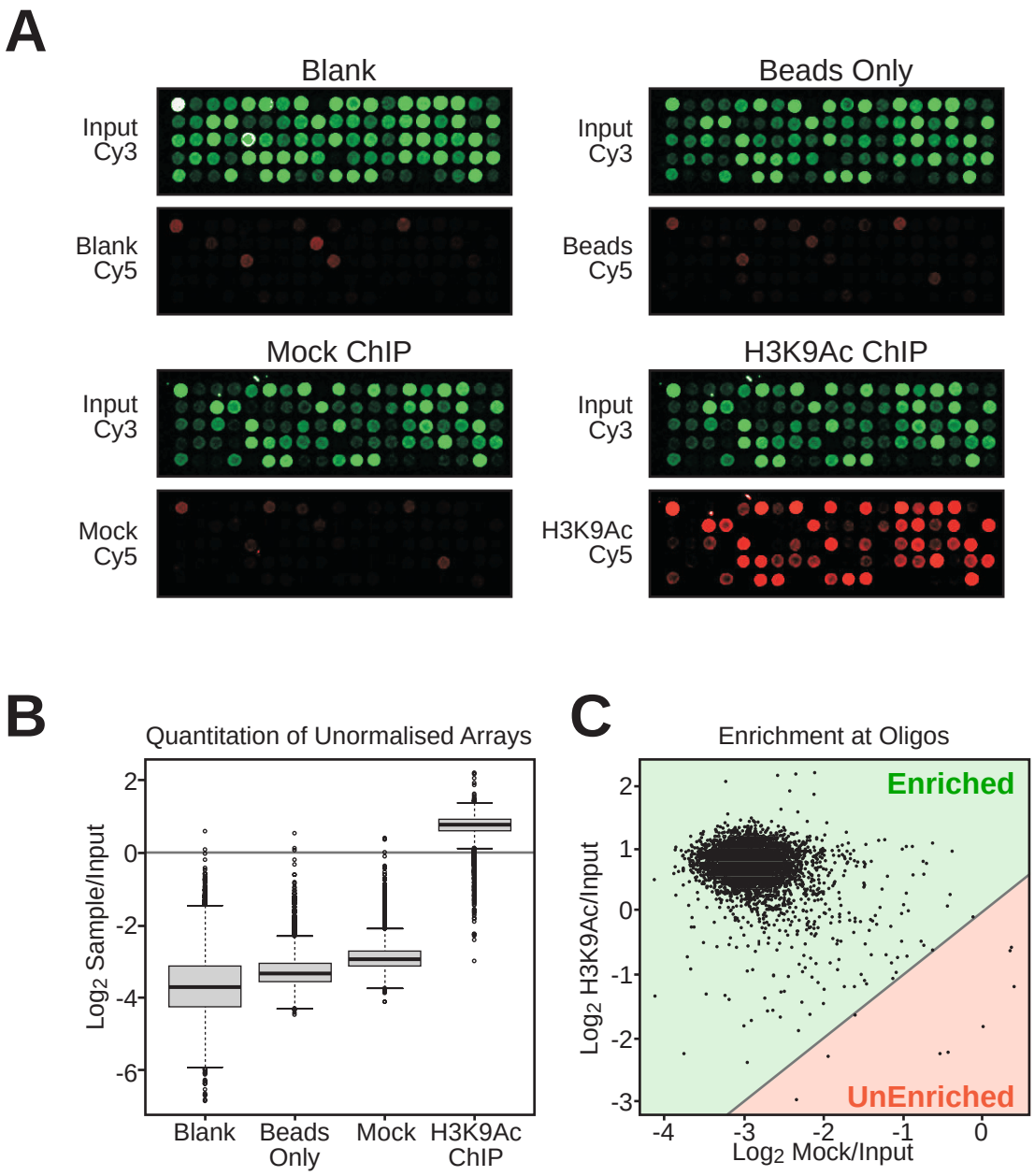


Figure 5.17 Analysis of N-ChIP on Microarrays

A. Images of Scanned Arrays. All Arrays were hybdridised with the same input sample and scanned at the same power and PMT Gain. Material was from 293 cells.

B. Quantitation of 293 ChIP Hybridistations. The same hybridisations shown in panel A were quantified and an average $\log_2$ sample/input value derived for each oligo. Boxplots of the overall distributions were then made without normalisation.

C. Enrichment of H3K9Ac Precipitations over Mock Precipitations. Log_2 H3K9Ac/Input was plotted against Log_2 Mock/Input. The placement of the values with regard to the grey line therefore represents the enrichment over mock for each oligo. Oligos in the green array are enriched relative to Mock precipitations and values in the red area are not (0.19%).

Most methodologies used to analyse ChIP-chip data also assume that probes on the array that are enriched over controls are fairly infrequent (~10%, (Buck and Lieb, 2004). The quantitation of signals from my control hybridisations suggested that the percentage of enriched probes in my experiment may be far higher (*Figure 5.17B*). Plotting $\log_2$ H3K9Ac/Input ratios against $\log_2$ Mock/Input showed that, nearly all probes were enriched over the mock precipitation (*Figure 5.17C*). The median enrichment was calculated as ~13x which compares favourably to previous estimates of between two and eightfold (Buck and Lieb, 2004) and references therein). It is not clear why such a high proportion of probes should be enriched over the mock precipitation, however, it has been noted that most sequences are enriched over mock controls during ChIP (Aparicio et al., 2004).

Another assumption made is that proteins are specifically bound or enriched at small sites in the genome and that any given site in the genome has one of two states, 'bound' or 'unbound'. By using arrays with a high resolution resulting from many closely tiled or overlapping probes, these binding sites can be identified as clusters of positive signals with enrichments lessening with distance from the actual binding site (Kim and Ren, 2006). However, it is unlikely that histone modifications can be considered in this manner as the amount of DNA precipitated is probably a function of the steady-state levels of modification at a particular nucleosome, which results from cycling between the modified and unmodified states (Metivier et al., 2003). ChIP-chip data for histone modifications probably therefore represents continuous rather than two-state data (Koch et al., 2007). Furthermore, I am using relatively low resolution arrays with roughly 1 probe every 800bp, equivalent to the size of chromatin fragment used, so I am unlikely to see peaks occurring across more than two probes. I therefore decided, initially at least, to analyse the data in a simple quantitative manner.

It is generally thought that histone acetylation is found at the promoters of active genes (Heintzman et al., 2007;Koch et al., 2007;Roh et al., 2005;Roh et al., 2006). To confirm that my protocol specifically assessed H3K9Ac patterns in cells, I determined whether I could detect enrichment of the mark at active gene promoters. I know that in HaCaT cells, the *S100A* genes are generally active whereas the *LCE* genes are inactive (*Chapter 3*). Therefore, by comparing the levels of H3K9Ac at

their respective promoters, I can assess the effectiveness of my protocol. I performed N-ChIP for H3K9Ac on HaCaT cells and hybridised the precipitated DNA to the EDC arrays twice in alternating dye orientations (ie two technical replicates of the hybridisation). The two hybridisations were normalised using a global lowess strategy and probes excluded on a dye swap basis before I calculated average probe $\log_2$ H3K9Ac/Input ratios (*Section 2.12.1*, p100). I then repeated the entire procedure and found that there was a high correlation between the $\log_2$ H3K9Ac/Input of the two biological replicates (*Figure 5.18A*). This suggests that specific patterns of acetylation are being measured because if the measured ratios reflected random noise we would expect no correlation.

The $\log_2$ values of probes in the region of the *S100A10* gene suggested that indeed *S100A* promoters were marked by H3K9Ac (*Figure 5.18B*). However, strong peaks of acetylation were also found just upstream and downstream of the genes promoter (*Figure 5.18B*, asterisks). Both peaks correspond roughly to what appeared to be conserved elements on the UCSC browser, and so may represent regulatory elements for the gene. By identifying probes found at all *S100A* promoters (ie lying -/+ 1Kb from transcriptional start sites as defined by the Refseq annotation, see boxed region *Figure 5.18A+B*), I found that the distribution of $\log_2$ H3K9Ac/Input of probes at *S100A* promoters was significantly shifted towards positive values when compared to all other probes on the array (*Figure 5.18C*, KS-test, p-value 2.264×10^{-07}) and was also above the level of experimental error as assessed by control self hybridisiations. When I also examined the probes located at the promoters of *LCE* genes I found the distribution was significantly different from that of the *S100A* promoters (*Figure 5.18C*) and did not show a shift towards positive values. This would suggest that compared to the promoters of active *S100A* genes, promoters of inactive *LCE* genes are depleted in H3K9Ac in HaCaT cells, suggesting that my protocol does indeed measure specific patterns of acetylation.

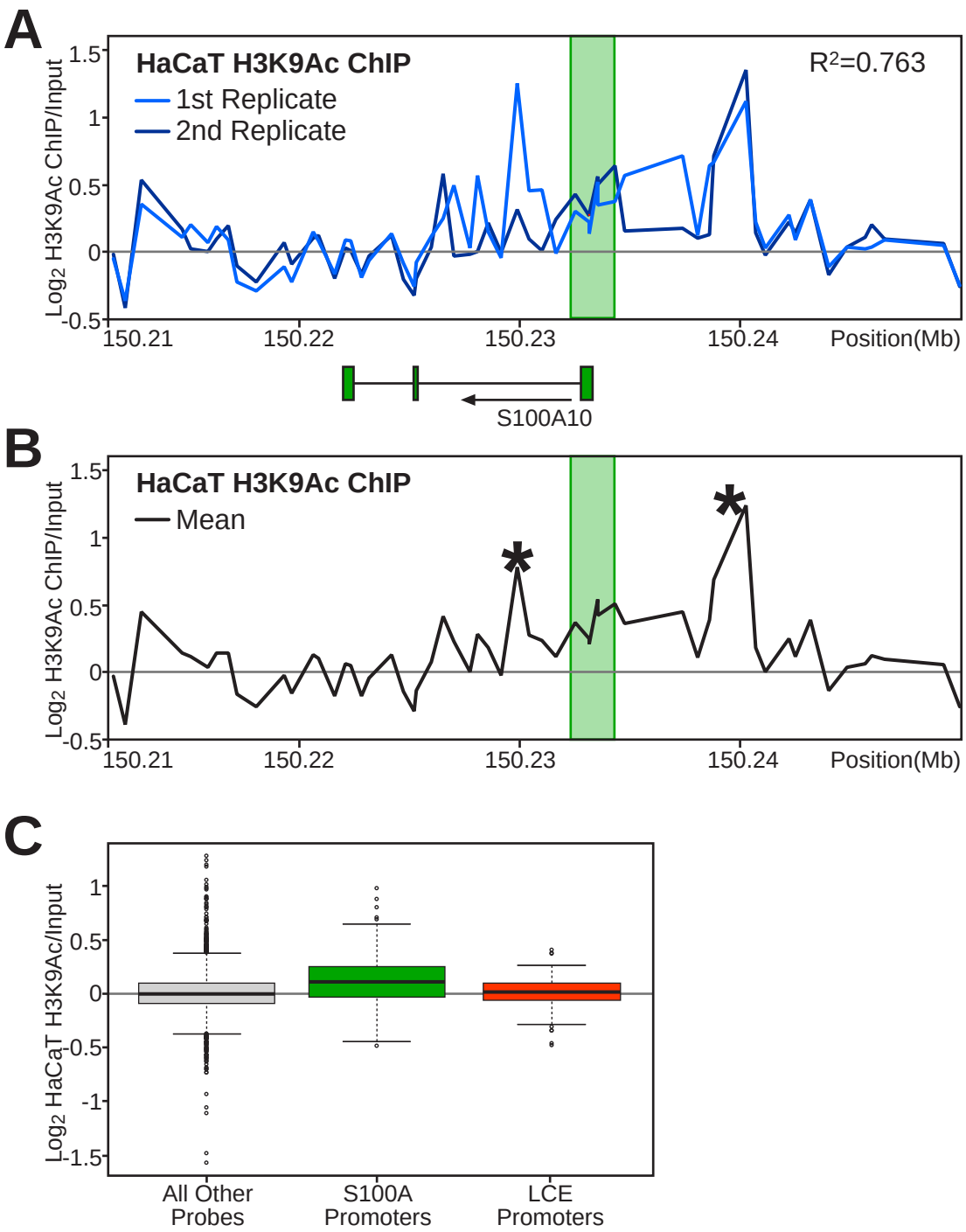


Figure 5.18 Assesment of H3K9Ac ChIP in HaCaT Cells

A. Plot of two biological replicates of HaCaT H3K9Ac ChIP in the region of the *S100A10* gene. The mean of two technical hybridisations of each replicate is shown. The position of the *S100A10* gene is indicated below the plot and its promoter shown by the green box (Refseq annotation). The Pearson's correlation (R^2) between the two replicates is also given.

B. As A. the mean of the two biological replicates is shown. The asterisks indicate peaks of acetylation lying adjacent to the promoter.

C. Boxplot showing the distribution of $\log_2$ H3K9Ac/Input values observed for oligos -/+ 1Kb from the transcriptional starts of *S100A* (Green) and *LCE* genes (Red). The distribution of all other probes on the array is also shown. All three distributions are significantly different (KS-test,p-value between S100A and LCE =3.32×10^{-04}).

5.10 Generation of ChIP Datasets

Having established that my ChIP approach can be used to assess the distribution of histone modifications across the entire EDC, I now wanted to examine the relationship between histone modifications and secondary chromatin structure. Previous experiments, performed *in vitro*, have suggested that histone acetylation may be capable of modulating the structure of 30nm chromatin fibres (Garcia-Ramirez et al., 1995;Tse et al., 1998). However, to date no study has examined the relationship between acetylated histones and secondary chromatin structure *in vivo*. Histones can be acetylated on a large number of different lysine residues including Histone H3 Lysines 9 and 14 and Histone H4 Lysines 5, 8, 12 and 16. It is known that the tail of histone H4 between residues 14 to 23 makes important internucleosomal contacts and its deletion inhibits the formation of compact nucleosomal arrays *in vitro* (Dorigo et al., 2003). In another study it was found that the acetylation of Lysine 16 alone inhibits formation of compact fibres *in vitro* (Shogren-Knaak et al., 2006). Therefore, I felt that this modification was a good candidate for the modulation of chromatin structure. Because I already had data on the distribution of H3K9Ac in HaCaT cells, I also decided to examine its relationship with secondary chromatin structure. Finally, I also wanted to examine the distribution of a repressive histone mark at the EDC. Since I was examining the pattern of H3K9Ac I decided to examine the distribution of H3K9 methylation which should show the opposite pattern to H3K9Ac. H3K9 can either be mono-, di- or tri-methylated, and it was thought at the time these experiments were performed, that H3K9me3 mainly marked constitutively heterochromatic regions whereas H3K9me2 can be found more widely in euchromatin (although H3K9me3 can also be found there, (Gilbert et al., 2003). Therefore, I elected to study the distribution of H3K9me2 across the EDC.

To generate data on these histone modifications, I performed ChIP on native chromatin from both HaCaT and 293 cells using commercial antibodies against H3K9Ac, H4K16Ac and H3K9me2. For each antibody and cell line, I independently performed the experiment twice and then each ChIP sample was labelled and hybridised to the EDC oligonucleotide tiling arrays twice. This means that I had

biologically repeated the experiment twice for each cell line and histone modification but that I had 4 separate hybridisations for each (ie two technical replicates of the hybridisation for each sample). I then normalised all hybridisations using a global lowess method, excluded probes based on the dye swap and calculated the mean $\log_2$ ChIP/Input value for each probe from the 4 hybridisations (*Section 2.12.1*, p100).

These experiments produced data on the pattern of histone modifications across the EDC for all 3 histone modifications (*Figures 5.19* and *5.20*). The data could be used to examine the association of modifications with different promoters and regulatory elements. However, this type of analysis has been performed by numerous other studies using far larger datasets and higher resolution systems (Barski et al., 2007;Heintzman et al., 2007;Koch et al., 2000;Roh et al., 2004). Therefore, I decided to concentrate on examining the relationship between histone modifications and higher levels of chromatin structure. However, initial analysis suggested that, as expected, the level of H3K9Ac and H3K9me2 were negatively correlated in both cell lines (data not shown). H4K16Ac was seen to be associated with some active promoters, but $\log_2$ ChIP/Input values for H4K16Ac and H3K9Ac were not highly correlated (data not shown). This would suggest that H4K16Ac and H3K9Ac have a rather different distribution across the EDC. In addition, the spread of the distribution of H4K16Ac $\log_2$ values was lower than that of H3K9Ac values, which may indicate a lower degree of enrichment during ChIP using the H4K16Ac antibody. In general, epitopes which are more abundant in the genome would be expected to give lower enrichments during ChIP than less abundant epitopes (Buck and Lieb, 2004). Therefore, H4K16Ac may be more abundant in the genome than H3K9Ac but this result could also easily be generated by a difference in the efficiencies of the two antibodies.

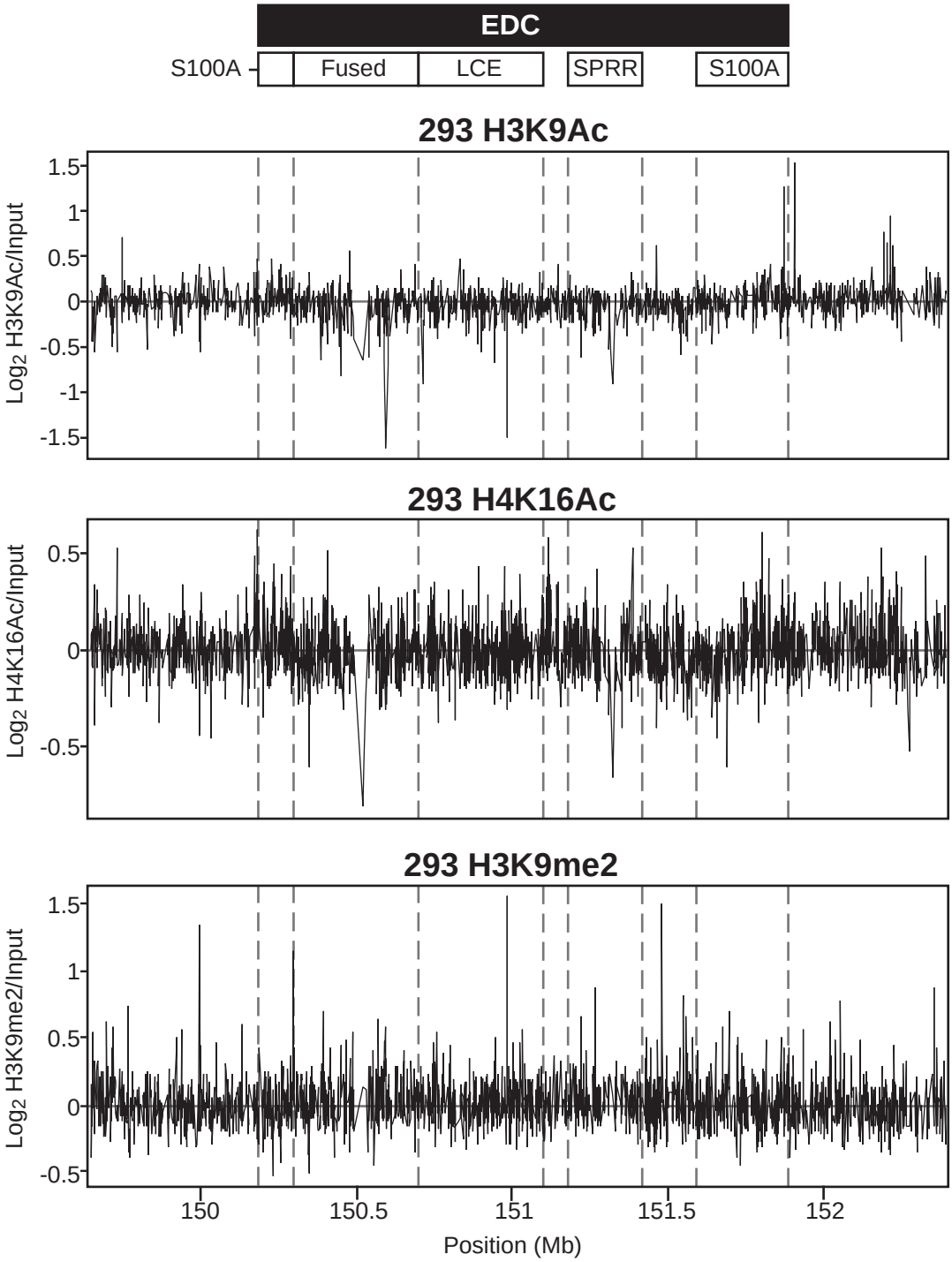


Figure 5.19 Histone Modifications Across the EDC in 293 Cells

ChIP was performed on MNase digested chromatin from 293 cells using anti-H3K9Ac, H4K16Ac or H3K9me2. Precipitated DNA was then hybridised to the EDC oligonucleotide tiling arrays and resulting $\log_2$ ChIP/Input ratios plotted by genomic position. Shown are the means of 4 hybridisations (2 biological replicates of the ChIP and 2 technical replicate hybridisations of each of these in oppossite dye orientations) after the exclusion of probes based on the dye swap. Genomic Positions are taken from NCBI build 36.

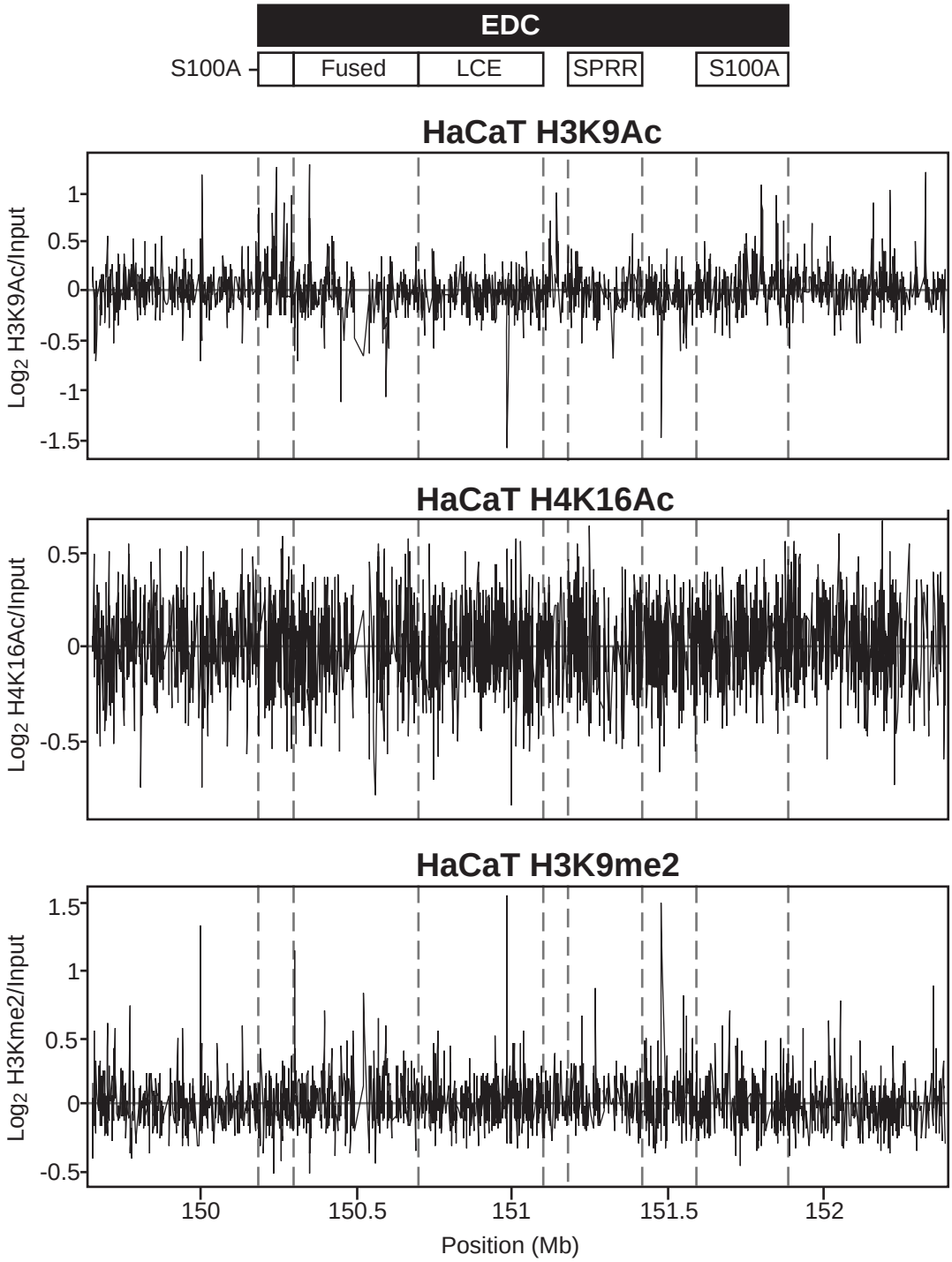


Figure 5.20 Histone Modifications Across the EDC in HaCaT Cells

ChIP was performed on MNase digested chromatin from HaCaT cells using anti-H3K9Ac, H4K16Ac or H3K9me2. Precipitated DNA was then hybridised to the EDC oligonucleotide tiling arrays and resulting $\log_2$ ChIP/Input ratios plotted by genomic position. Shown are the means of 4 hybridisations (2 biological replicates of the ChIP and 2 technical replicate hybridisations of each of these in opposite dye orientations) after the exclusion of probes based on the dye swap. Genomic Positions are taken from NCBI build 36.

5.11 Is there Correlation Between Primary and Secondary Chromatin Structure?

An *in vitro* study has suggested that the formation of compact chromatin fibres is inhibited if nucleosome arrays are made from H4K16Ac marked nucleosomes (Shogren-Knaak et al., 2006). If this was an important mechanism *in vivo*, I would expect to see evidence of this in my data in the form of a correlation between secondary chromatin structure and H4K16Ac levels. The simplest relationship would be that the modification locally affected the structure of the chromatin fibre and therefore we might expect that it would be possible to detect a correlation between $\log_2$ H4K16Ac/Input values and $\log_2$ open/input values. However, I found only a small correlation between H4K16Ac levels and secondary chromatin structure in both HaCaT and 293 cells (*Table 5.3*). Therefore, if H4K16Ac does alter chromatin structure *in vivo*, it is clearly not the single most important factor at the EDC. I found similarly low correlations between the datasets for H3K9Ac, H3K9me2 and secondary chromatin structure (*Table 5.3*) suggesting that none of these single modifications are they major determinants of secondary chromatin structure at the EDC.

It is possible that acetylation does affect chromatin structure, but that no single modification is important for this effect and instead there is a relationship between the overall level of acetylation and secondary chromatin structure. To comprehensively analyse this, the distribution of all acetylations would have to be analysed. However, we can perhaps assume that in a region with a high overall level of acetylation, both H3K9 and H4K16 would be acetylated. Therefore, I calculated the correlation of $\log_2$ H3K9Ac/Input + $\log_2$ H4K16Ac/Input with secondary chromatin structure. This correlation was lower than for the two individual modifications suggesting that total local acetylation levels also did not correlate in a simple manner with secondary chromatin structure (*Table 5.3*).

Table 5.3

Correlations Between Chromatin Structure Array Datasets

Log_2 ChIP/input values of all probes on the array were correlated with log_2 open/input values. All correlation values are Kendall's correlations.

Comparison	Correlation 293	Correlation HaCaT
H3K9Ac	-0.140	0.045
H4K16Ac	0.068	0.070
H3K9diMe	0.106	0.136
H3K9Ac + H4K16Ac	-0.051	0.068

I wondered whether, although these individual histone modifications did not correlate with secondary chromatin structure on small genomic scales, perhaps the level of modifications across a chromatin domain was correlated with its structure. However, it is difficult to determine which scale this might occur at, so rather than assume a particular scale, I decided to scan scales from 1Kb to 100Kb. I calculated mean log_2 values across the region using a 1Kb sliding window for both the H3K9Ac ChIP data and secondary structure data. The Kendall's correlation of the two profiles was then calculated at the 1Kb scale. I then repeated the process using sliding windows of 1.5Kb, 2Kb, 2.5Kb, etc up to 100Kb to comprehensively analyse the correlation between histone modifications and secondary chromatin structure at genomic scales between 1 and 100Kb. The level of correlation was then plotted against the genomic scale (ie sliding window size) for both HaCaT and 293 cells (*Figure 5.21*). In HaCaT cells the correlation between the two datasets increased to a peak at ~60.5Kb but in 293 cells no such peak was observed. However, the correlation between H3K9Ac and secondary structure in HaCaT cells, even at the peak, was of a low level and I found it not to be significant by random permutation of the data (*Table 5.4*). Therefore, it is unlikely to represent a biologically relevant correlation. I similarly examined the correlation between the other histone modifications (H4K16Ac and H3K9me2) and secondary chromatin structure at

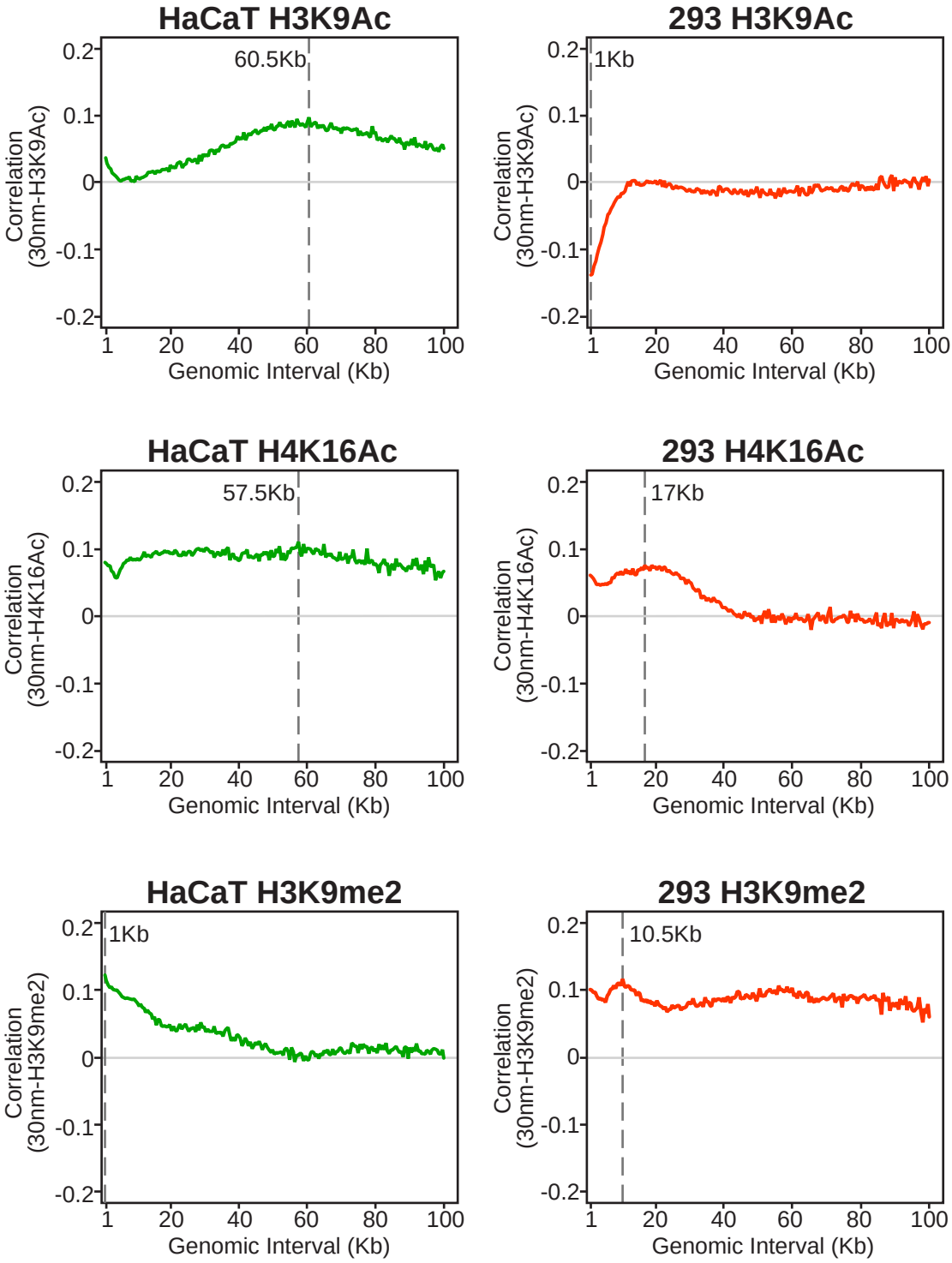


Figure 5.21 Correlation Between Primary and Secondary Levels of Chromatin Structure

Sliding windows were used to calculate mean open/input and ChIP/input values for window sizes between 1Kb and 100Kb across the region of the EDC. For each window size, the Kendall's correlation was calculated between the datasets and a graph of correlation against window size plotted. Windows were moved across the locus in increments 1/10th of their size. In all cases the indicated histone modifications were compared to secondary chromatin structure for the two cell lines. The dashed line on each plot, indicates the interval size at which the maximum correlation was found.

intervals up to 100Kb (*Figure 5.21*, and *Table 5.4*). H4K16Ac showed a weak correlation with secondary structure just above that expected for random data in HaCaT cells but not in 293 cells. Therefore, I conclude that across scales of 1 to 100Kb, the levels of single histone modifications do not greatly correlate to secondary chromatin structure and that if histone modification levels do determine the organisation of chromatin at this scale, it is through the interaction of different modifications rather than a single important modification.

Table 5.4

Table of Tested Correlation Peaks and P-values

By analysing the correlation between ChIP datasets (log_2 ChIP/input ratios) and secondary chromatin structure (log_2 open/input ratios) over genomic intervals between 1Kb and 100Kb, I identified the genomic interval at which the two datasets showed the greatest correlation (*Figure 5.21*). I then assessed the biological relevance of this correlation by randomly permutating the datasets 1000 times and calculating the Kendall's correlation at the genomic interval in question for each random permutation. The p-value of the correlation was then the percentage of randomised datasets which had a correlation greater than or equal to the observed correlation. Correlations at small genomic intervals were not analysed due to the limitations of computing power.

Cell	Modification	Interval (Kb)	Kendall's Correlation	p-value
HaCaT	H3K9Ac	60.5Kb	0.097	0.124
293	H3K9Ac	1	-0.138	-
HaCaT	H4K16Ac	57.5	0.110	0.083
293	H4K16Ac	17	0.076	0.046
HaCaT	H3Kme2	1	0.122	-
293	H3Kme2	10.5	0.116	-

5.12 Histone Modifications also Form Large Scale Domains at the EDC

Although I have not observed a correlation between histone modifications and secondary chromatin structure at scales less than 100Kb, histone modification

patterns have previously been seen to correlate with the chromosomal banding pattern (Barski et al., 2007;Jeppesen, 1997;Roh et al., 2005) as has secondary chromatin structure (Gilbert et al., 2004). This suggests a correlation exists between histone modification patterns and secondary chromatin structure at large scales (>1Mb). I cannot easily assess such large genomic scales due to the size of my array (~2.5Mb). However, I examined whether histone modifications are also organised into large scale domains corresponding to those found in secondary structure at the 250Kb scale (*Section 5.4*). Therefore, I analysed H3K9Ac, H4K16Ac and H3K9me2 patterns using a 250Kb sliding window and compared the resulting profiles to that of secondary chromatin structure analysed in the same manner (*Figures 5.22* and *5.23*).

Firstly, it was apparent at this scale that the histone modifications themselves had highly correlated profiles. H3K9Ac and H4K16Ac followed similar profiles in both cell lines and there was correlation between them (Kendall's 0.507 in 293s and 0.312 in HaCaTs). H3K9Ac and H3K9me2 were negatively correlated in both cell lines (Kendall's -0.667 in 293s and -0.356 in HaCaTs). The modifications were organised into large scale domains of a similar nature in both cell lines. High levels of acetylation were found at the edges of the EDC in both cell lines and H3K9me2 was generally spread throughout the locus. A recent large scales study had examined H3K9me2 distributions throughout the genome (using ChIP followed by mass sequencing), and using their data I found that the profile of H3K9me2 at the inactive EDC was similar in both 293 and human T-cells (Barski et al., 2007), Kendall's = 0.630).

Since the large scale organisation of the histone modifications was similar between the cell lines, we might expect that it would not correlate to secondary chromatin structure which we know differs between the two cell lines. When the region as a whole was analysed, this was indeed the case (data not shown). When the EDC alone was analysed, higher correlations were seen between H3K9Ac, H4K16Ac and secondary chromatin structure in HaCaT cells but these were not above the level of random association (*Table 5.5*). Therefore, like secondary structure, histone modifications appear to be organised into domains at the EDC. However, these domains do not correlate greatly with similar domains observed in

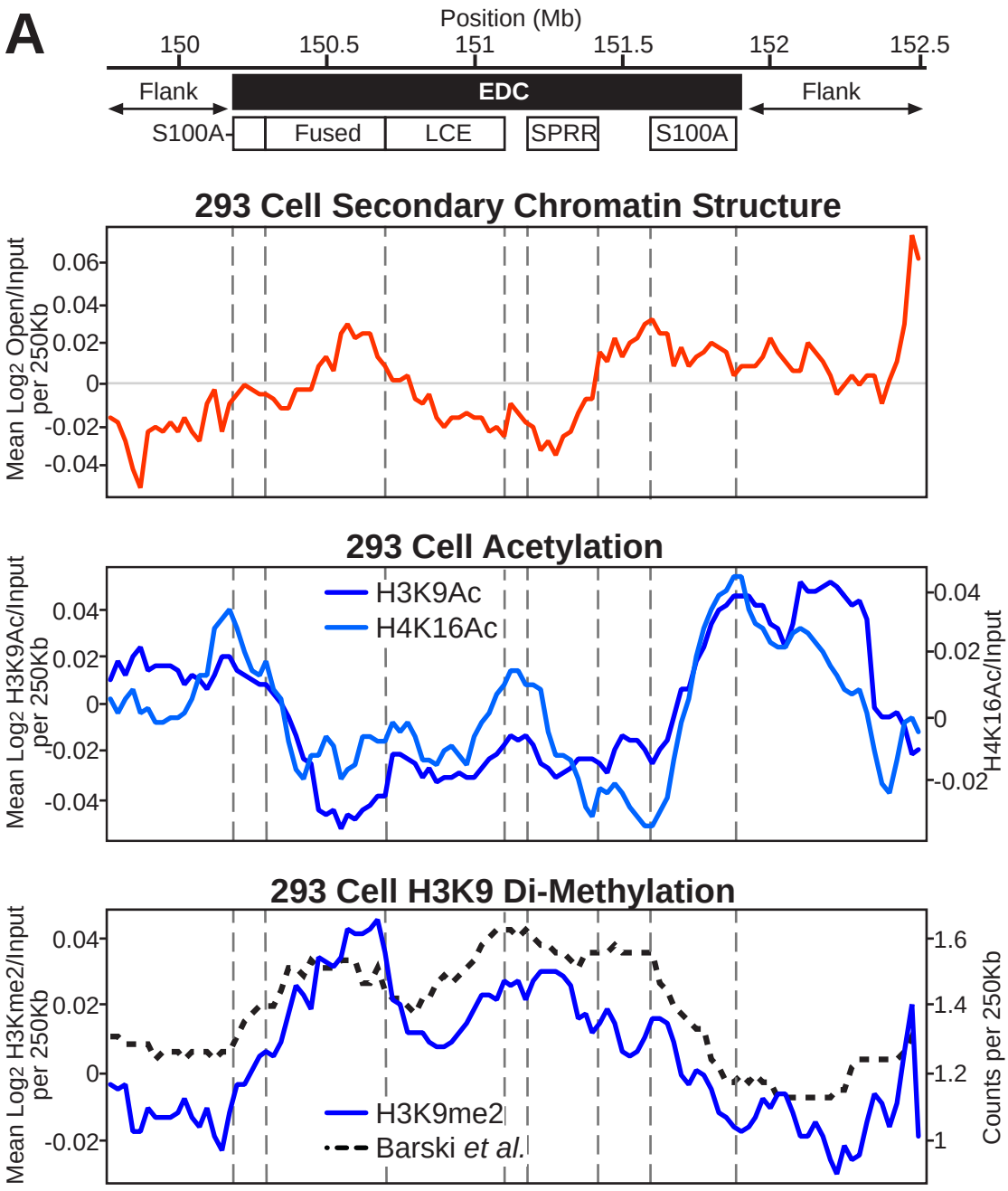


Figure 5.22 Correlation between Secondary Chromatin Structure and Histone Modifications at Large Scales in 293 Cells

A. Plot of chromatin fibre structure against genomic position across the EDC. A 250Kb sliding window was passed along the region in 25Kb increments to calculate the mean $\log_2$ open/input ratio. The position of the EDC and its gene clusters are indicated. Genomic positions are taken from NCBI build 36.

B. Acetylation at the EDC. As for A. but data on histone acteylation from ChIP was used instead. The left hand axis is for H3K9Ac and the right for H4K16Ac.

C. H3K9me2 at the EDC. As for A. but data on H3K9me2 from ChIP was used instead. The left hand axis is for ChIP data from 293 cells (blue). The dashed line shows the data from a recent study which examined many histone modifications in human T-cells (Barski *et al.*, 2007) and is plotted on the right hand axis. This data is derived from mass sequencing and quantified in tag counts. The two are highly correlated, Kendall's = 0.630.

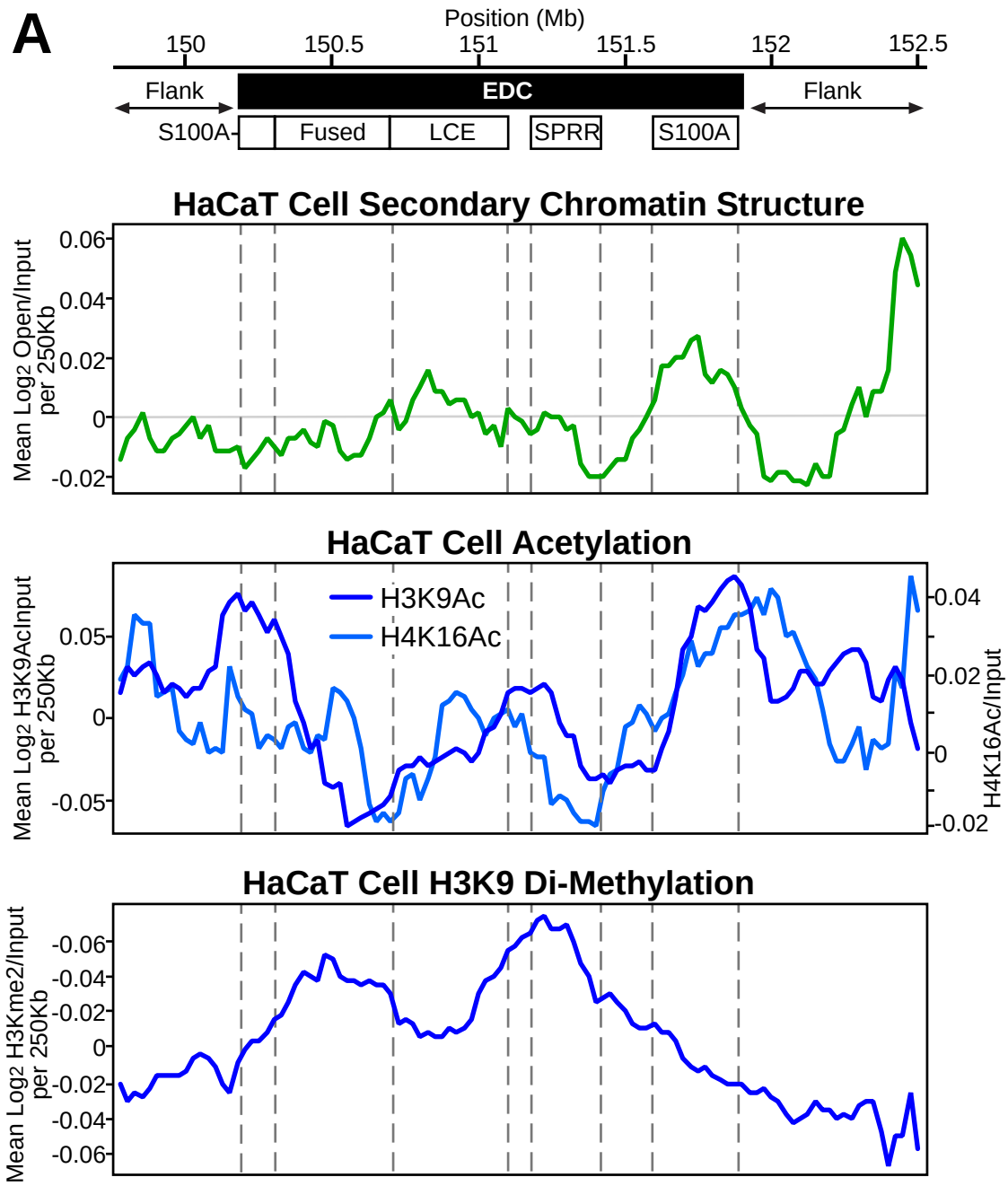


Figure 5.23 Correlation between Secondary Chromatin Structure and Histone Modifications at Large Scales in HaCaT Cells

A. Plot of chromatin fibre structure against genomic position across the EDC. A 250Kb sliding window was passed along the region in 25Kb increments to calculate the mean $\log_2$ open/input ratio. The position of the EDC and its gene clusters are indicated. Genomic positions are taken from NCBI build 36.

B. Acetylation at the EDC. As for A. but data on histone acteylation from ChIP was used instead. The left hand axis is for H3K9Ac and the right for H4K16Ac.

C. H3K9me2 at the EDC. As for A. but data on H3K9me2 from ChIP was used instead.

the profile of secondary structure across the region. The organisation of the histone modification domains is similar between 3 cell lines (at least for H3K9me2). In support of these findings, another recent study has observed the organisation of histone modifications into domains that are similar between cell types, but did not examine secondary chromatin structure (Thurman et al., 2007).

Table 5.5

Correlations Between Chromatin Structure Array Datasets at the 250Kb Scale Within the EDC

Mean $\log_2$ ChIP/input values were calculated using a sliding window of 250Kb and correlations between the secondary structure and histone modification profiles calculated. All correlation values are Kendall's correlations. A p-value was assigned to each correlation by random permutation analysis were the values of each probe within the region were replaced by a sample from the whole array and the analysis repeated 1000 times. The p-value of the correlation was then the percentage of the randomly generated correlations greater to or equal to the observed value.

Comparison	Correlation 293	p-value 293	Correlation HaCaT	p-value HaCaT
H3K9Ac	0.060	0.396	0.179	0.195
H4K16Ac	-0.163	0.199	0.277	0.079
H3K9diMe	-0.113	0.278	-0.293	0.078

5.13 Nucleosomes are Not Generally Depleted from the Promoter Regions of Active Genes

A final aspect of chromatin structure that I wanted to examine was nucleosome occupancy. It has been suggested in a recent ChIP-chip study, that histones are generally depleted from the region around the transcriptional start sites of active genes (Koch et al., 2007). This study used crosslinking ChIP of Histones H3 and H2B to assess nucleosome occupancy and the observed depletion could be due to epitope occlusion (as discussed in *Section 5.8*). To test this more directly, I isolated

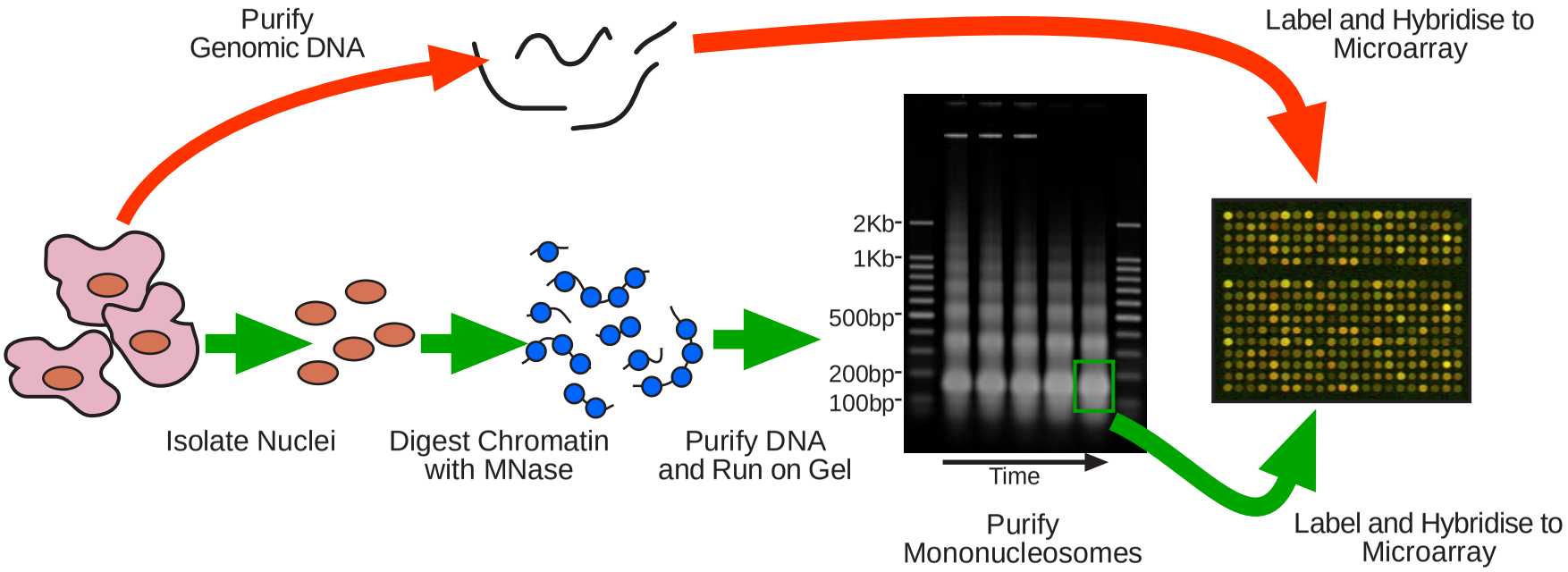


Figure 5.24 Assement of Nucleosome Position at the EDC

Nuclei were isolated from the cell lines and digested with MNase. Aliquots were removed throughout the digestion and the DNA purified and run on agarose gels. Form this a suitable timepoint was selected were most of the chromatin had been digested. This timepoint was then run on a separate gel and the mononucleosome band excised and the DNA purified. This was then labelled and co-hybridised with genomic DNA on microarrays to assess nucleosomal position.

nuclei from both HaCaT and 293 cells and digested them with MNase. Aliquots of nuclei were removed along a time course, the DNA purified and resolved on a gel. I selected a timepoint where no high molecular weight DNA was visible on the gel and the vast majority of nucleosomes were present as monomers, dimers and trimers (*Figure 5.24*). I then resolved only this fraction on a separate gel and cut out the band corresponding to the mononucleosomes. After purifying the DNA, it was labelled and hybridised to the EDC oligonucleotide tiling arrays with purified genomic DNA. Two technical replicates of the hybridisation were performed in alternating dye orientations and the entire experiment was repeated twice for each cell line (ie two biological replicates and 4 hybridisations per cell line total).

On these hybridisations, a high $\log_2$ mononucleosome (mono)/Genomic ratio signified a nucleosome was strongly positioned at that probe and a low ratio signified a nucleosome was generally absent. The correlation between the two biological replicates of the experiment in HaCaT cells was high (Pearson's $R^2 = 0.705$) suggesting a consistent pattern of nucleosome positioning did exist. In addition, comparison between the distribution of average $\log_2$ mono/genomic ratios observed for four hybridisations from HaCaT cells and 293 cells was significantly wider than the resulting average $\log_2$ ratios of four control self hybridisations (*Figure 5.25A*, KS-tests $p<2.2x10^{-16}$). If nucleosomes were not occupying preferential positions we would expect to see a similar distribution compared to control hybridisation. Log_2 values for each probe were highly correlated between 293 and HaCaT cells suggesting that nucleosome positions may be conserved between cell lines (*Figure 5.25B*).

Examination of the region surrounding the active *S100A10* in HaCaT cells suggested that nucleosomes were not depleted from its promoter region (*Figure 5.25C*). In fact a peak of positive $\log_2$ values appeared to be positioned very close to the transcriptional start site of the gene. The distribution of $\log_2$ mono/genomic ratios observed for probes -/+ 2Kb from the transcriptional start sites of *S100A* genes was not significantly different from that of all other probes (KS-test, p=0.058, *Figure 5.25D*). Therefore, no general nucleosome loss appears to occur around active *S100A* gene promoters.

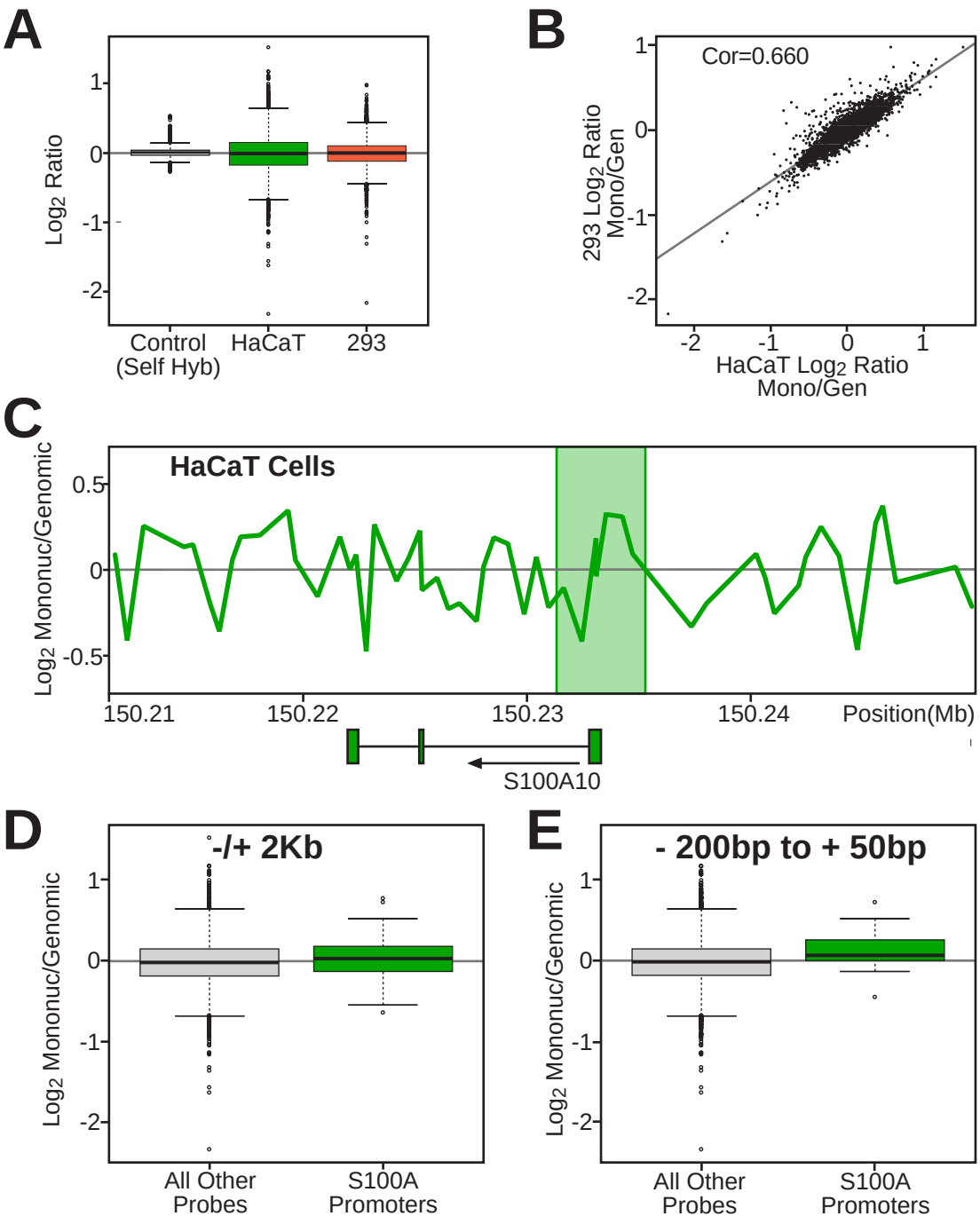


Figure 5.25 Promoters of Active S100A Genes are not depleted in Nucleosomes

A. Boxplot of expermental and control hybridisations. The distributions from HaCaT and 293 cells are significantly different from that of the control (KS-tests)

B. Correlation (Kendall's) between $\log_2$ Mono/Genomic Ratios for HaCaT and 293 cells.

C. Plot of $\log_2$ mono/gen across *S100A10* region in HaCaTs with the promoter highilighted.

D. Boxplots showing distributions of mononucleosome/genomic rations for *S100A* gene promoters versus the rest of the array in HaCaT cells. Promoters defined as -/+ 2Kb from transcriptional starts (Refseq UCSC March 2006).

E. As for D but for an interval of -200 to +50bp from transcriptional starts.

A recent large scale ChIP study has suggested that nucleosome depletion at active promoters is confined to a single nucleosome, positioned at around -200 to +50bp from the transcription start site (Barski et al., 2007). However, they did not specifically examine nucleosome occupancy and arrived at this conclusion from an observed dip in H3K4me abundance within this region. I examined the distribution of log values for probes found within this smaller interval and surprisingly found that their distribution was shifted towards +ve values (*Figure 5.25E*). This shift was not quite significant, probably because of the low numbers of probes around the promoters of S100A genes (14 probes, KS-test, p=0.051) but it was above experimental error as determined from self hybridisations. I also examined the same interval for all genes within the EDC and also found a shift but despite the increase in the number of probes, this was less significant than for the S100A promoters alone (data not shown, 36 probes, KS-test, p=0.141). A similar shift was also seen for nucleosomes from 293 cells at the *S100A* promoters and this was also not significant but above experimental error (KS-test, p=0.094). Therefore, clearly there is a difference in the two studies. This may simply mean that nucleosomes are present at some types of active promoters and the fact that the previous study was on such a large scale meant that they were not detected. Secondly, it may be because Barski *et al.*, did not directly examine nucleosome occupancy and so derived their data by performing ChIP for histone modifications. This may mean that nuclesomes so positioned are not H3K4 methylated or that they are tightly bound by factors which obscure the epitopes for ChIP. It has been seen in another study that a nucleosome is strongly positioned at the transcriptional start site of the chicken adult *β-globin* gene (Davey et al., 1995).

5.14 Discussion

In this chapter, I have examined the chromatin structure of the EDC at the primary and secondary levels, ie the placement and modification of nucleosomes and their arrangement into chromatin fibres. I first examined bulk chromatin structure in both cell lines using sucrose gradients and showed that they are similar, before combining sucrose gradients with microarrays to show that, overall, the EDC does not have a

different secondary chromatin structure from its surrounding genomic regions. This finding was in agreement with my analysis of interphase chromatin compaction in the region (*Chapter 4*) and I also find that the interphase chromatin compaction and 30nm chromatin fibre structures are correlated at the region, in agreement with a previous study from our laboratory (Gilbert et al., 2004). When analysed in greater detail, the EDC appears to be organised into large domains of both secondary chromatin structure and histone modifications. At the active EDC in HaCaT cells, large scale secondary organisation correlates to the gene density in the region but the inactive EDC in 293 cells has a different structure that may reflect LINE density. Large scale secondary structure does not reflect GC content at the EDC in either cell line. I also find that these large scale domains may be composed of smaller chromatin domains, which appear to have an irregular size. Although histone modifications are also organised into domains in the EDC, these are fairly constant between cell lines and no single histone modification significantly correlates to secondary chromatin organisation at genomic scales from ~1-250Kb. By isolating mononucleosomes and hybridising them to microarrays, I have also examined nucleosomal positioning in both cell lines. The patterns of the two cell lines were highly correlated, but contrary to recent reports I found no evidence of a general depletion of nucleosomes around the promoters of active genes (Koch et al., 2007) or of the loss of a single nucleosome over the transcriptional start site (Barski et al., 2007).

Chapter 6: Conclusions and Future Perspectives

Although many studies have examined the relationship between chromatin structure and transcription either across the whole genome or at a particular locus, few studies have looked at the relationship between different levels of chromatin structure. Here I have used a variety of techniques to study the chromatin structure of the human EDC and to begin to describe how the different levels of chromatin structure relate to each other. In addition, my study is, at present, one of the few examining the overall regulation of this important locus and provides us with insights into the regulation of the EDC *in vivo*.

6.1 Does the EDC Represent a Separate Chromatin Domain?

Our current knowledge suggests that gene clusters behave as independent chromatin domains from their surrounding genomic regions and it would seem reasonable to suppose that the EDC is regulated in a similar manner (see *Figure 1.6*, p48). Data from previous studies suggested that the inactive locus is marked by repressive histone modifications. Furthermore, a study of nuclear localisation suggested that the active locus moved out of its chromosome territory, something that is often accompanied by a decondensation of chromatin structure (Williams et al., 2002). However, these studies had not examined the secondary or tertiary structure of the locus in detail and it is unclear whether a ~2Mb supercluster should behave in a manner analogous to a ~100Kb gene cluster.

To examine tertiary chromatin structure, I used a set of genomic clone pairs to perform FISH and measure the interphase compaction of the region. This assay has previously been employed to measure differences in regional compaction in the human genome, albeit on the level of chromosome bands (Yokota et al., 1997). Allowing for the difference in general interphase compaction I found in HaCaT cells,

the profiles of the HaCaT and control cell lines appear very similar (*Figure 4.8* p147). Importantly, their profiles do not suggest that the EDC behaves as a domain that is independent its surrounding. The majority of the EDC has a similar compaction to the surrounding regions and although the distal portion of the EDC is decondensed relative to the rest of the genomic region, it appears constitutively so.

Despite these similarities, I have showed that the profile of compaction in the region is subtly different between the control cell lines and HaCaT cells. Furthermore, by comparing differences in compaction across the region to differences in gene expression (*Chapter 3*), I find a surprising negative correlation between transcriptional activity and interphase compaction. The reasons for this are unclear, but we know that at active genes, regulatory elements can loop together to facilitate transcription (Tolhuis et al., 2002). Across a large genomic region (for example the 500Kb intervals examined here), such looping may be sufficient to cause a cytologically measurable degree of compaction. If this were true, it might be predicted that the relationship would break down if compaction was measured across smaller genomic intervals.

When I examined secondary chromatin organisation in the region using a biophysical method, I found a fairly direct relationship between it and tertiary chromatin organisation (*Figure 5.6*, p166), in agreement with a previous study from our laboratory (Gilbert et al., 2004). In addition, like the analysis of tertiary chromatin structure, this experiment suggests that the EDC does not represent a single large chromatin domain that is condensed when inactive and decondensed when active.

Examination of the region in detail suggests that it is in fact made up of multiple domains of chromatin structure, for example the distal *S100A* cluster is present in a domain of relatively open chromatin in both cell lines (*Figure 5.7*, p168). However, in these two cell lines the organisation of these domains differs. In HaCaT cells domains enriched in open chromatin fibres are associated with gene-rich parts of the locus, whereas in 293 cells they associate, albeit more weakly, with regions rich in LINE elements. Taken together these results suggest that the EDC does not represent a single large chromatin domain and is instead made up of smaller domains which upon activation of the locus undergo rearrangement, but not general

decondensation. More generally, it suggests the genome is made up of large chromatin domains, represented by chromatin bands (Gilbert et al., 2004;Goetze et al., 2007) which are themselves are made up of smaller chromatin domains, like those seen here. These domains are large enough to be assessed by FISH using pairs of probes (van den Engh et al., 1992). The probe pairs used in *Chapter 4* were selected arbitrarily, without prior knowledge of the organisation of the locus and cross the boundaries of these domains. More appropriate pairs could be selected and used to gain independent verification that a series of chromatin domains exist at both the active and inactive EDC.

I have so far assumed that the arrangement of chromatin domains found in HaCaT cells represents the EDC in its active state and that in 293 the inactive state. However, this may not be the case and if I investigated other cell lines with either active or inactive loci, I may find a different arrangement of domains. Therefore, one set of further experiments would be to examine the chromatin structure of other cell lines to deduce which features of chromatin organisation are consistently found at the active and inactive EDC.

6.2 The Relationship Between Histone Modifications and Higher Levels of Chromatin Structure

Given that histone modifications, like secondary chromatin structure, appear organised at the level of the chromosome band (Barski et al., 2007;Jeppesen, 1997;Roh et al., 2005) and that smaller scale chromatin domains of histone modifications have been described (Thurman et al., 2007), I expected the domains of secondary structure to correspond to large domains of histone modifications. However, when I examined the distribution of histone modifications by ChIP, I found that although they were organised into large domains, these did not correspond to the secondary chromatin structure (*Figures 5.22* and *5.23*, p202 and 203). The EDC itself is relatively depleted in acetylated histones and enriched in H3K9me2 modified histones compared to the flanking regions. However, levels of the

modifications were not constant across the locus and it is composed of smaller domains, for example the central portion of the EDC around the LCE cluster appears relatively enriched in acetylated histones relative to the rest of the EDC. These domains are of a similar size to those recently described as part of the ENCODE project (Thurman et al., 2007).

If both histone modifications and secondary chromatin structure are organised into large scale domains, why are the two sets of domains not correlated? Assuming that the primary function of most histone modifications is to regulate chromatin structure, my results suggest that a simple histone code does not exist. That is to say histone modifications do not act in a simple combinatorial manner to directly alter chromatin structure. One study in yeast has inferred a simple combinatorial code for the majority of Histone H4 acetylation using mutants and expression microarrays but they did not examine other levels of chromatin structure (Dion et al., 2005). The yeast genome is also far simpler than mammalian genomes, and perhaps a more complex histone code has evolved for a more complex genome (Garcia et al., 2007).

My data suggests that if histone modifications provide a template for setting up secondary chromatin structure in the cell, they probably do so through the interaction of a large number of different modifications. Therefore, local secondary chromatin structure is determined by the action of different chromatin remodelling enzymes that are recruited by individual histone modifications. Similarly to a recent study of the ENCODE regions, I find that the arrangement of histone modification domains are similar between cell lines (Thurman et al., 2007). I would propose therefore, that if histone modifications do determine secondary chromatin structure, the observed differences in secondary chromatin structure at the EDC are caused by relatively small shifts in the distributions of a number of different histone modifications. In addition, other factors such as histone variants probably also play a role, further complicating the situation.

If histone modifications determine secondary chromatin structure in a complicated, combinatorial manner, it will be a laborious process to examine the distribution of all modifications and establish how this occurs. It may be easier to study the results of their actions. We know that in the majority of the genome, modifications occur in discrete areas seen as ‘peaks’ when examined by ChIP-chip

(Bernstein et al., 2005). A subset of modifications such as H3K4 methylation are found in regions that are highly coincident with sites of DNase hypersensitivity (Birney et al., 2007). The study of the chicken *β-globin* locus has shown that hypersensitive sites can create discontinuities in chromatin fibres which cause them to sediment more slowly in a sucrose gradient (Caplan et al., 1987). I have examined secondary structure by quantifying the relative abundance of chromatin fibres which behave in this manner. Therefore, it might be expected that, if I examined the distribution of DNase hypersensitive sites across the region, the 'openness' of a particular part of locus might correlate to the number of DNase hypersensitive sites or discontinuities within that region. This could also be examined by quantifying the number of histone modification peaks within a region but we know that different combinations of modifications are associated with different categories of hypersensitive sites and therefore multiple modifications would have to be examined if this approach was taken (Birney et al., 2007;Heintzman et al., 2007).

6.3 The Nature of Chromatin Domains Observed at the EDC

Although I have described the existence of chromatin domains within the EDC, I have not determined the nature of these domains. It is clear that the larger domains of ~100-500Kb in size are themselves made up of smaller domains that appear to be less than ~100Kb (*Figure 5.10*, p175) and, at least in HaCaT cells, these smaller domains are of a fairly regular size (*Section 5.7*, p174). Histone modifications also appear to be organised into similar domains (data not shown) but the two sets of domains do not correlated (*Section 5.11*, p197).

We understand little of chromatin organisation above the 30nm fibre but a number of different models have been proposed. One model is that interphase chromosomes are organised into loops that are attached to a backbone like structure and each loop is a chromatin domain (*Figure 1.2*, p25 and (Sachs et al., 1995;Yokota et al., 1995). This is a similar model to that proposed for metaphase chromosome organisation (Boy de la Tour and Laemmli, 1988). However, an implication of this

model is that because domains exist on different loops, they are independent from each other. At the EDC it appears that the reorganisation of the locus results in a rearrangement of the adjacent genomic regions also. This implies that adjacent domains are not independent and that the reorganisation of the regions flanking the EDC is required to compensate for the rearrangement of the locus itself. Such an arrangement might suggest that higher order structure is generated by the further folding of the 30nm fibre into thicker chromatin fibres (Belmont and Bruce, 1994). One could imagine that the reorganisation of part of such a structure would have an impact on the structure of an adjacent part of the higher order fibre. This is consistent with the fact that measurements of interphase compaction suggest that at most genomic regions the chromatin fibre is folded at levels above the 30nm fibre (Chambeyron and Bickmore, 2004;Gilbert et al., 2004;Yokota et al., 1997). My data is not inconsistent with the model of giant loops proposed for interphase chromosomes, because these loops were proposed to be ~2-3Mb in size, but it suggests that different chromatin domains of a non-looplike nature exist within these loops (Sachs et al., 1995;Yokota et al., 1995). If the higher order fibres were arranged into loops that are attached at either end, this may actually impose more constraints on the conformation of the higher order chromatin fibre of the loop.

It is possible that the smaller domains represent the pattern caused by the purification of preferentially digested fragments with retarded mobility in the sucrose gradient. Therefore, further work using independent methods should be used to determine their nature. We know that different parts of the genome have different digestion properties when analysed using nucleases. For example, mouse minor satellite which is more compact digests more slowly with MNase than bulk chromatin (Gilbert et al., 2007) and we might expect a more open region to digest more rapidly. By adapting the method used in *Section 5.13* (p204) to examine nucleosomal positioning, it may be possible to examine the digestion properties of chromatin domains within the EDC. If the chromatin is digested more briefly with MNase, the portion of the genome present in the mononucleosome fraction is likely to come mainly from more easily digested, open domains. More compact domains are likely to be refractory to digestion and depleted from this fraction. If the analysis of this chromatin fraction on the EDC oligonucleotide arrays reveals a domain

pattern similar to that of the sedimentation analysis, this would suggest that domains of relatively more open and closed chromatin do exist in the genome.

A further method of analysing the nature of the different chromatin domains is to use the 3C technique (Dekker, 2007). Using this technique, an inverse correlation is found between the frequency of interaction of two genomic locations and the genomic distance between them. In a less compact region, the frequency of interaction falls off more quickly than in a more compact region. Such an approach could be used to examine the structural properties of the domains identified in this study.

Further analysis of these chromatin domains should also attempt to define their boundaries in a more precise manner. Recently the combination of wavelet analysis and a Hidden Markov model was used to define domains of histone modifications, replication timing and expression in the ENCODE project (Thurman et al., 2007). Wavelet analysis is a method, like power spectrum analyses, that has been applied in physics and engineering and has recently been applied to biological problems (Lio, 2003). The main difference in the Wavelet transformation when compared to the Fourier transformation is the inclusion of a spatial component. This means that the analysis is capable of detecting a wavelike structure in the data even with the presence of sharp peaks and discontinuities, which we may expect from the analysis or irregularly sized chromatin domains. Another approach may be to adapt the Smith-Waterman algorithm to define regions enriched in open or compact structures (Price et al., 2005).

It is possible that specific genomic elements exist at the boundaries between these chromatin domains, and by more accurately defining where these boundaries are, it may be possible to identify such elements. The nature of these elements may also tell us something about the nature of chromatin domains. For example if they are matrix attachment sites, it may suggest that the domains represent loops attached at either end to the nuclear matrix (Mirkovitch et al., 1984). If the domains described here are the result of higher order folding of the chromatin fibre, we may expect that such elements would not exist at the boundaries of the domains.

One particular protein that may be involved in setting up the different domains of the EDC is CCCTC-binding factor (CTCF). It is a ubiquitously

expressed protein which plays an important role in setting up chromatin insulators blocking the interaction of enchancers with promoters (Bell et al., 1999). Recently, CTCF binding in human T-cells was examined using ChIP and mass sequencing and found to be present at the boundries between domains of different histone modifications (Barski et al., 2007). Another study used ChIP-chip to examine CTCF binding in human primary fibroblasts and a hematopoietic cell line (Kim et al., 2007). 67% of the binding sites in the two cell lines were found to overlap, indicating that the protein may have a role in setting up the basic domain organisation of the genome. On average in the fibroblast genome it was found that pairs of CTCF binding sites defined domains of 2.5 genes in length but, a number of larger domains were identified corresponding to clusters of co-ordinately regulated genes, including the *LCE* cluster. By examining the publicly available data from these studies, it is apparent that CTCF binding sites are present close to the boundries of many of the gene clusters within the EDC (*Figure 6.1*). Therefore, it would interesting to examine where CTCF binds in my model cell lines and how these binding sites correspond to the chromatin domains I have described.

6.4 A Program of Gene Expression at the EDC?

Many years ago, it was proposed that the EDC exists to facilitate the co-ordinate regulation of genes encoding constituents (Mischke, 1998). However, my study is the first comprehensive study of the chromatin organisation of the region in any organism. Therefore, despite using cell lines as models, the data can be used to gain insight into how chromatin structure may regulate the EDC *in vivo*. The first stage of examining the mechanisms governing the EDC is to describe the pattern of gene expression they control. The literature suggests that during epidermal differentiation, gene expression may start on the edges of the EDC and proceed into the centre of the locus over time (*Figure 3.4*, p119). The results of my study suggest that such a pattern may occur in HaCaT cells.

However, HaCaTs are a cultured cell line and further experiments are required to determine if this pattern of expression exists *in vivo*. The expression

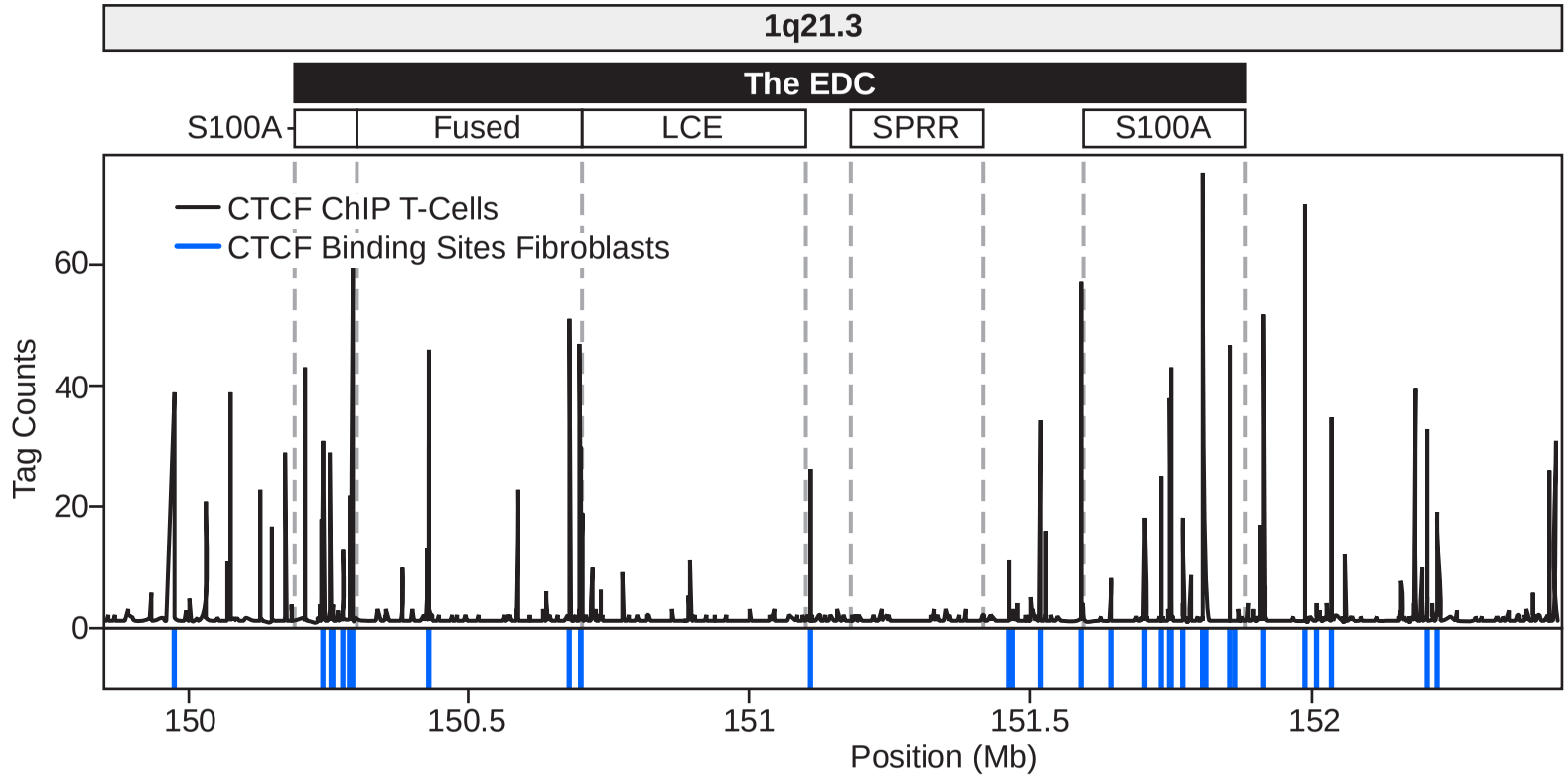


Figure 6.1 CTCF Binding Within the EDC

Map of the EDC showing CTCF Binding in two different cell lines. The graph in black shows data from the mass sequencing of a CTCF ChIP experiment in human $CD4^+$ T-cells (Barski et al., 2007). Below this in blue are CTCF binding sites in human IMR90 primary fibroblasts (Kim et al., 2007). Both data sets are publically available for viewing in the UCSC browser and I plotted the graph using this data. The data of Barski et al., is annotated in NCBI 36. However, that of Kim et al., is annotated in NCBI 35 so I had to map it onto NCBI 36. In this region, the sole apparent difference between the two assemblies is that NCBI 36 genomic coordinates are offset from those in NCBI 35 by +1453551bp. Therefore this value was added to the midpoints of the annotated binding sites before they were plotted.

pattern of primary keratinocytes and *in vitro* skin models have already been examined using microarrays (Gazel et al., 2003;Mehul et al., 2004) but these studies have not examined how gene expression alters with differentiation. Another study has compared gene expression in basal and differentiating keratinocytes using microarrays (Radoja et al., 2006). Their findings support the hypothesis that gene expression begins in the *S100A* clusters bounding the EDC and that the later clusters are active in differentiating cell but, the study focused on differential gene expression genome wide rather than the activation of the EDC. In addition, the differentiating cell population used may contain a mix of cells at different stages, therefore, the data does not allow us to determine whether the *Fused*, *SPRR* and *LCE* clusters become activated together or sequentially. Therefore, I propose that the differentiation of primary keratinocytes could be used to analyse how the EDC is activated as differentiation proceeds. This allows a number of different timepoints to be taken and the temporal activation of different parts of the cluster to be examined.

To examine the locus as a whole, microarrays should be used but I would probably adopt a different experimental approach to that used in *Chapter 3*. The use of a reference RNA makes it difficult to interpret the results of the experiments in a quantitative manner. Proprietary microarray systems such as that supplied by Affymetrix are more accurate and enable the quantitative measurement of absolute transcript levels (Dalma-Weiszhausz et al., 2006). However such systems are often prohibitively expensive and an alternative approach is to use genomic DNA as a reference sample (Williams et al., 2006). By doing so it would be possible for me to use the EDC oligonucleotide arrays to generate quantitive data on gene expression at the EDC. These arrays are also tiling arrays and so this approach has the additional advantage of allowing both genic and non-genic transcription at the EDC to be examined. It has recently come to light through the efforts of the ENCODE project and other studies that a far larger proportion of the human genome than previously thought is transcribed (Birney et al., 2007;Kapranov et al., 2002). It would be interesting to examine how the proposed pattern of expression relates to non-genic transcription as well as the sequence of activation of the genes themselves.

However, examining gene expression in this manner quantifies steady state RNA levels rather than transcription itself. The fact that a certain degree of post

transcriptional regulation occurs during epidermal differentiation (*Section 3.6*, p126) suggests that mRNA levels may not necessarily represent transcriptional activity levels. One approach to examine transcription more directly is to perform ChIP for the RNA polymerase complex and so quantify its level at different parts of the locus (Sandoval et al., 2004). Another approach is to use nuclear run-on assays which, thus far have not been combined with genomic microarrays. However, it has been shown that the use of biotin labelled nucleotides can allow RNA synthesised in a nuclear run on assay to be isolated and analysed by Quantitative PCR (Patrone et al., 2000). It should, therefore, be possible to amplify this RNA and analyse its distribution across the EDC using microarrays.

6.5 HaCaT Cells as a Model of EDC Function

My analysis suggests that HaCaT cells represent an early stage of epidermal differentiation which is consistent with their ability to be differentiated *in vitro* (Boukamp et al., 1988;Paramio et al., 1998). They express mainly the genes associated with early epidermal differentiation, for example the *S100A* genes. As mentioned in *Chapter 1*, the role of the *S100A* genes in the epidermis remains slightly mysterious. However, my expression data in *Chapter 3* shows that they are far more highly expressed in HaCaT cells than in the control cell lines, reinforcing that they probably play and important role in this tissue. This is explains why the might be encoded within the EDC which is otherwise composed of highly specialised genes that are only expressed during epidermal differentiation.

If HaCaT cells represent an early stage of epidermal differentiation, the chromatin organisation of the active EDC described in this study is likely to correspond to that found in either the basal or lower spinous layers of the epidermis (*Figure 1.5*, p40). Therefore, it is possible that this organisation represents a state that primed or poised for further reorganisation as later gene clusters become expressed in subsequent differentiation stages. This will most probably mean that the H3K9me2 marks found in the central portion of the EDC in HaCaT cells will be removed and replaced with more active marks. However, it is unclear whether this

will accompanied by a further rearrangement of secondary or tertiary chromatin structure.

The best way to investigate the changes in chromatin organisation that occur *in vivo* is to perform FISH on tissue sections made from skin biopsies. Using appropriately selected probes, this system could be used to show that features of tertiary chromatin structure found in HaCaT cells also occur *in vivo*, for example the location of different chromatin domains. Such an analysis will also allow us to examine how the arrangement of chromatin domains changes as differentiation proceeds. FISH could also be used to examine the overall tertiary organisation of the locus in sections if probes composed of pooled BACs are used (Goetze et al., 2007).

One observation in my study that contradicts a previous study, is the fact that I found that the EDC occupied relatively similar positions for the EDC in HaCaT and 575 cells (*Section 4.2*, p138). It has previously been reported that, in primary keratinocytes the EDC is positioned externally to the chromosome territory, and in lymphoblastoid cells it is positioned within or at the edge of the chromosome 1 territory (Williams et al., 2002). The difference in the two studies could be attributed to differences between HaCaTs and primary keratinocytes but if I analyse my data in a similar manner to this previous study, I find that the nuclear localisation of the EDC in HaCaT cells is actually very similar to that reported for primary keratinocytes (data not shown). Therefore, the difference between the two studies appears to be in the localisation of the EDC in two different lymphoblastoid cell lines. In support of my observations however, I have examined the nuclear position of the locus in another independent control cell line, 293 cells and find that in these cells the EDC is also similarly positioned to that in HaCaT cells.

Finally, in the course of this study, I have generated a large quantity of data on primary chromatin structure at the EDC. As the EDC remains little studied, we do not know much about regulatory elements that are present within it. One study has used comparative genomics to identify elements at the *SPRR* cluster (Martin et al., 2004). However, the ENCODE project has suggested that regulatory elements are not necessarily evolutionarily conserved but are epigenetically marked (Birney et al., 2007). Therefore, one approach to identify novel elements at the EDC would be to analyse the data by clustering to identify probes that are marked by H3K9me2 in

293 cells and by acetylation in HaCaT cells (Gollub and Sherlock, 2006). Further, studies of the locus, for example the large scale identification of DNase hypersensitive sites, can be used to identify such sites with more certainty.

6.6 Long Term Goals

Although this study has provided us with insight into the organisation of chromatin structure and the relationship between different levels of chromatin structure, this work and the further experiments detailed above remain rather descriptive in nature. To demonstrate that these features of chromatin structure have a role in regulating the locus, it will be necessary to show that perturbation of chromatin structure results in the misregulation of the EDC. However, in order to achieve this, we will need to identify factors capable of regulating the locus and its chromatin structure. Identified binding sites of CTCF make it a strong candidate for setting up the patterns of domains at the locus. Therefore, it would be interesting to interfere with the protein (eg by RNAi) to see what effect this has on the regulation and chromatin organisation of the locus. A more subtle experiment would be to interfere with individual binding sites within the locus to examine if they contribute to the overall organisation of the region.

Bibliography

Abbott,D.W., V.S.Ivanova, X.Wang, W.M.Bonner, and J.Ausio. 2001. Characterization of the stability and folding of H2A.Z chromatin particles: implications for transcriptional activation. *J. Biol. Chem.* **276**: 41945-41949.

Ahmad,K. and S.Henikoff. 2002. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**: 1191-1200.

Aho,H., M.Schwemmer, D.Tessman, D.Murphy, G.Mattei, W.Engel, and I.M.Adham. 1996. Isolation, expression, and chromosomal localization of the human mitochondrial capsule selenoprotein gene (MCSP). *Genomics* **32**: 184-190.

Alibardi,L. 2002. Immunocytochemical localization of keratins, associated proteins and uptake of histidine in the epidermis of fish and amphibians. *Acta Histochem.* **104**: 297-310.

Alibardi,L. 2003. Adaptation to the land: The skin of reptiles in comparison to that of amphibians and endotherm amniotes. *J. Exp. Zoolog. B Mol. Dev. Evol.* **298**: 12-41.

Alibardi,L., E.Spisni, and M.Toni. 2003. Presence of putative histidine-rich proteins in the amphibian epidermis. *J. Exp. Zoolog. A Comp Exp. Biol.* **297**: 105-117.

Alibardi,L. and M.Toni. 2004. Immuno-cross reactivity of transglutaminase and cornification marker proteins in the epidermis of vertebrates suggests common processes of soft cornification across species. *J. Exp. Zoolog. B Mol. Dev. Evol.* **302**: 526-549.

Allan,J., G.J.Cowling, N.Harborne, P.Cattini, R.Craigie, and H.Gould. 1981. Regulation of the higher-order structure of chromatin by histones H1 and H5. *J. Cell Biol.* **90**: 279-288.

Allan,J., P.G.Hartman, C.Crane-Robinson, and F.X.Aviles. 1980. The structure of histone H1 and its location in chromatin. *Nature* **288**: 675-679.

Allan,J., D.C.Rau, N.Harborne, and H.Gould. 1984. Higher order structure in a short repeat length chromatin. *J. Cell Biol.* **98**: 1320-1327.

Allfrey,V.G., R.Faulkner, and A.E.Mirsky. 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A* **51**: 786-794.

Allfrey,V.G., V.C.Littau, and A.E.Mirsky. 1963. On the role of of histones in regulation ribonucleic acid synthesis in the cell nucleus. *Proc. Natl. Acad. Sci. U. S. A* **49**: 414-421.

Allis,C.D. and J.C.Wiggins. 1984. Histone rearrangements accompany nuclear differentiation and dedifferentiation in Tetrahymena. *Dev. Biol.* **101**: 282-294.

Almouzni,G., S.Khochbin, S.Dimitrov, and A.P.Wolffe. 1994. Histone acetylation influences both gene expression and development of Xenopus laevis. *Dev. Biol.* **165**: 654-669.

Aparicio,O., J.V.Geisberg, and K.Struhl. 2004. Chromatin Immunoprecipitation for determining the Association of Proteins with Specific Genomic sequences In Vivo. *Current Protocols In Cell Biology* **Supplement 23**: 17.7.1-17.7.23.

Arents,G., R.W.Burlingame, B.C.Wang, W.E.Love, and E.N.Moudrianakis. 1991. The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. U. S. A* **88**: 10148-10152.

Ashe,H.L., J.Monks, M.Wijgerde, P.Fraser, and N.J.Proudfoot. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* **11**: 2494-2509.

Aviles,F.J., G.E.Chapman, G.G.Kneale, C.Crane-Robinson, and E.M.Bradbury. 1978. The conformation of histone H5. Isolation and characterisation of the globular segment. *Eur. J. Biochem.* **88**: 363-371.

Bannister,A.J., R.Schneider, and T.Kouzarides. 2002. Histone methylation: dynamic or static? *Cell* **109**: 801-806.

Bannister,A.J., P.Zegerman, J.F.Partridge, E.A.Miska, J.O.Thomas, R.C.Allshire, and T.Kouzarides. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**: 120-124.

Barratt,M.J., C.A.Hazzalin, E.Cano, and L.C.Mahadevan. 1994. Mitogen-stimulated phosphorylation of histone H3 is targeted to a small hyperacetylation-sensitive fraction. *Proc. Natl. Acad. Sci. U. S. A* **91**: 4781-4785.

Barski,A., S.Cuddapah, K.Cui, T.Y.Roh, D.E.Schones, Z.Wang, G.Wei, I.Chepelev, and K.Zhao. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823-837.

Batada,N.N. and L.D.Hurst. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* **39**: 945-949.

Bell,A.C., A.G.West, and G.Felsenfeld. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387-396.

Belmont,A.S. and K.Bruce. 1994. Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J. Cell Biol.* **127**: 287-302.

Benoit,S., A.Toksoy, M.Ahlmann, M.Schmidt, C.Sunderkotter, D.Foell, M.Pasparakis, J.Roth, and M.Goebeler. 2006. Elevated serum levels of calcium-binding S100 proteins A8 and A9 reflect disease activity and abnormal differentiation of keratinocytes in psoriasis. *Br. J. Dermatol.* **155**: 62-66.

Berger,J.A., S.K.Mitra, and J.Astola. 2003. Power spectrum analysis for DNA sequences. *Proceedings of the Seventh International Symposium on Signal Processing and its Applications* **2**: 29-32.

Bernstein,B.E., E.L.Humphrey, R.L.Erlich, R.Schneider, P.Bouman, J.S.Liu, T.Kouzarides, and S.L.Schreiber. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A* **99**: 8695-8700.

Bernstein,B.E., M.Kamal, K.Lindblad-Toh, S.Bekiranov, D.K.Bailey, D.J.Huebert, S.McMahon, E.K.Karlsson, E.J.Kulbokas, III, T.R.Gingeras, S.L.Schreiber, and E.S.Lander. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**: 169-181.

Bernstein,B.E., T.S.Mikkelsen, X.Xie, M.Kamal, D.J.Huebert, J.Cuff, B.Fry, A.Meissner, M.Wernig, K.Plath, R.Jaenisch, A.Wagschal, R.Feil, S.L.Schreiber, and E.S.Lander. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.

Bhattacharya,S., C.G.Bunick, and W.J.Chazin. 2004. Target selectivity in EF-hand calcium binding proteins. *Biochim. Biophys. Acta* **1742**: 69-79.

Birney,E., J.A.Stamatoyannopoulos, A.Dutta, R.Guigo, T.R.Gingeras, E.H.Margulies, Z.Weng, M.Snyder, E.T.Dermitzakis, R.E.Thurman, M.S.Kuehn, C.M.Taylor, S.Neph, C.M.Koch, S.Asthana, A.Malhotra, I.Adzhubei, J.A.Greenbaum, R.M.Andrews, P.Flicek, P.J.Boyle, H.Cao, N.P.Carter, G.K.Clelland, S.Davis, N.Day, P.Dhami, S.C.Dillon, M.O.Dorschner, H.Fiegler, P.G.Giresi, J.Goldy, M.Hawrylycz, A.Haydock, R.Humbert, K.D.James, B.E.Johnson, E.M.Johnson, T.T.Frum, E.R.Rosenzweig, N.Karnani, K.Lee, G.C.Lefebvre, P.A.Navas, F.Neri, S.C.Parker, P.J.Sabo, R.Sandstrom, A.Shafer, D.Vetrie, M.Weaver, S.Wilcox, M.Yu, F.S.Collins, J.Dekker, J.D.Lieb, T.D.Tullius, G.E.Crawford, S.Sunyaev, W.S.Noble, I.Dunham, F.Denoeud, A.Reymond, P.Kapranov, J.Rozowsky, D.Zheng, R.Castelo, A.Frankish, J.Harrow, S.Ghosh, A.Sandelin, I.L.Hofacker, R.Baertsch, D.Keefe, S.Dike, J.Cheng, H.A.Hirsch, E.A.Sekinger, J.Lagarde, J.F.Abril, A.Shahab, C.Flamm, C.Fried, J.Hackermuller, J.Hertel, M.Lindemeyer, K.Missal, A.Tanzer, S.Washietl, J.Korbel, O.Emanuelsson, J.S.Pedersen, N.Holroyd, R.Taylor, D.Swarbreck, N.Matthews, M.C.Dickson, D.J.Thomas, M.T.Weirauch, J.Gilbert, J.Drenkow, I.Bell, X.Zhao, K.G.Srinivasan, W.K.Sung, H.S.Ooi, K.P.Chiu, S.Foissac, T.Alioto, M.Brent, L.Pachter, M.L.Tress, A.Valencia, S.W.Choo, C.Y.Choo, C.Ucla, C.Manzano, C.Wyss, E.Cheung, T.G.Clark, J.B.Brown, M.Ganesh, S.Patel, H.Tammana, J.Chrast, C.N.Henrichsen,

C.Kai, J.Kawai, U.Nagalakshmi, J.Wu, Z.Lian, J.Lian, P.Newburger, X.Zhang, P.Bickel, J.S.Mattick, P.Carninci, Y.Hayashizaki, S.Weissman, T.Hubbard, R.M.Myers, J.Rogers, P.F.Stadler, T.M.Lowe, C.L.Wei, Y.Ruan, K.Struhl, M.Gerstein, S.E.Antonarakis, Y.Fu, E.D.Green, U.Karaoz, A.Siepel, J.Taylor, L.A.Liefer, K.A.Wetterstrand, P.J.Good, E.A.Feingold, M.S.Guyer, G.M.Cooper, G.Asimenos, C.N.Dewey, M.Hou, S.Nikolaev, J.I.Montoya-Burgos, A.Loytynoja, S.Whelan, F.Pardi, T.Massingham, H.Huang, N.R.Zhang, I.Holmes, J.C.Mullikin, A.Ureta-Vidal, B.Paten, M.Seringhaus, D.Church, K.Rosenbloom, W.J.Kent, E.A.Stone, S.Batzoglou, N.Goldman, R.C.Hardison, D.Haussler, W.Miller, A.Sidow, N.D.Trinklein, Z.D.Zhang, L.Barrera, R.Stuart, D.C.King, A.Ameur, S.Enroth, M.C.Bieda, J.Kim, A.A.Bhinge, N.Jiang, J.Liu, F.Yao, V.B.Vega, C.W.Lee, P.Ng, A.Shahab, A.Yang, Z.Moqtaderi, Z.Zhu, X.Xu, S.Squazzo, M.J.Oberley, D.Inman, M.A.Singer, T.A.Richmond, K.J.Munn, A.Rada-Iglesias, O.Wallerman, J.Komorowski, J.C.Fowler, P.Couttet, A.W.Bruce, O.M.Dovey, P.D.Ellis, C.F.Langford, D.A.Nix, G.Euskirchen, S.Hartman, A.E.Urban, P.Kraus, C.S.Van, N.Heintzman, T.H.Kim, K.Wang, C.Qu, G.Hon, R.Luna, C.K.Glass, M.G.Rosenfeld, S.F.Aldred, S.J.Cooper, A.Halees, J.M.Lin, H.P.Shulha, X.Zhang, M.Xu, J.N.Haidar, Y.Yu, Y.Ruan, V.R.Iyer, R.D.Green, C.Wadelius, P.J.Farnham, B.Ren, R.A.Harte, A.S.Hinrichs, H.Trumbower, and H.Clawson. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.

Black,B.E., D.R.Foltz, S.Chakravarthy, K.Luger, V.L.Woods, Jr., and D.W.Cleveland. 2004. Structural determinants for generating centromeric chromatin. *Nature* **430**: 578-582.

Bode,J., M.M.Gomez-Lira, and H.Schroter. 1983. Nucleosomal particles open as the histone core becomes hyperacetylated. *Eur. J. Biochem.* **130**: 437-445.

Bolzer,A., G.Kreth, I.Solovei, D.Koehler, K.Saracoglu, C.Fauth, S.Muller, R.Eils, C.Cremer, M.R.Speicher, and T.Cremer. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS. Biol.* **3**: e157.

Bonaldi,T., A.Imhof, and J.T.Regula. 2004. A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics.* **4**: 1382-1396.

Bonaldi,T., G.Langst, R.Strohner, P.B.Becker, and M.E.Bianchi. 2002. The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. *EMBO J.* **21**: 6865-6873.

Boni,R., G.Burg, A.Doguoglu, E.C.Ilg, B.W.Schafer, B.Muller, and C.W.Heizmann. 1997. Immunohistochemical localization of the Ca2+ binding S100 proteins in normal human skin and melanocytic lesions. *Br. J. Dermatol.* **137**: 39-43.

Borun,T.W., K.Ajiro, A.Zweidler, T.W.Dolby, and R.E.Stephens. 1977. Studies of human histone messenger RNA. II. The resolution of fractions containing individual human histone messenger RNA species. *J. Biol. Chem.* **252**: 173-180.

Borun,T.W., F.Gabrielli, K.Ajiro, A.Zweidler, and C.Baglioni. 1975. Further evidence of transcriptional and translational control of histone messenger RNA during the HeLa S3 cycle. *Cell* **4**: 59-67.

Borun,T.W., D.Pearson, and W.K.Paik. 1972. Studies of histone methylation during the HeLa S-3 cell cycle. *J. Biol. Chem.* **247**: 4288-4298.

Boukamp,P., R.T.Petrussevska, D.Breitkreutz, J.Hornung, A.Markham, and N.E.Fusenig. 1988. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **106**: 761-771.

Boulikas,T., J.M.Wiseman, and W.T.Garrard. 1980. Points of contact between histone H1 and the histone octamer. *Proc. Natl. Acad. Sci. U. S. A* **77**: 127-131.

Boutanaev,A.M., L.M.Mikhaylova, and D.I.Nurminsky. 2005. The pattern of chromosome folding in interphase is outlined by the linear gene density profile. *Mol. Cell Biol.* **25**: 8379-8386.

Bowcock,A.M. and W.O.Cookson. 2004. The genetics of psoriasis, psoriatic arthritis and atopic dermatitis. *Hum. Mol. Genet.* **13 Spec No 1**: R43-R55.

Boy de la Tour,E. and U.K.Laemmli. 1988. The metaphase scaffold is helically folded: sister chromatids have predominantly opposite helical handedness. *Cell* **55**: 937-944.

Boyle,S., S.Gilchrist, J.M.Bridger, N.L.Mahy, J.A.Ellis, and W.A.Bickmore. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**: 211-219.

Branco,M.R. and A.Pombo. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS. Biol.* **4**: e138.

Broome,A.M., D.Ryan, and R.L.Eckert. 2003. S100 protein subcellular localization during epidermal differentiation and psoriasis. *J. Histochem. Cytochem.* **51**: 675-685.

Brown,D.T., T.Izard, and T.Misteli. 2006a. Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo. *Nat. Struct. Mol. Biol.* **13**: 250-255.

Brown,J.M., J.Leach, J.E.Reittie, A.Atzberger, J.Lee-Prudhoe, W.G.Wood, D.R.Higgs, F.J.Iborra, and V.J.Buckle. 2006b. Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell Biol.* **172**: 177-187.

Brown,K.E., S.Amoils, J.M.Horn, V.J.Buckle, D.R.Higgs, M.Merkenschlager, and A.G.Fisher. 2001. Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells. *Nat. Cell Biol.* **3**: 602-606.

Brown,S.W. 1966. Heterochromatin. *Science* **151**: 417-425.

Brownell,J.E., J.Zhou, T.Ranalli, R.Kobayashi, D.G.Edmondson, S.Y.Roth, and C.D.Allis. 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**: 843-851.

Bruce,K., F.A.Myers, E.Mantouvalou, P.Lefevre, I.Greaves, C.Bonifer, D.J.Tremethick, A.W.Thorne, and C.Crane-Robinson. 2005. The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res.* **33**: 5633-5639.

Bubulya,P.A. and D.L.Spector. 2004. "On the move"ments of nuclear components in living cells. *Exp. Cell Res.* **296**: 4-11.

Buck,M.J. and J.D.Lieb. 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**: 349-360.

Bulger,M., D.Schubeler, M.A.Bender, J.Hamilton, C.M.Farrell, R.C.Hardison, and M.Groudine. 2003. A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol. Cell Biol.* **23**: 5234-5244.

Burton,D.R., M.J.Butler, J.E.Hyde, D.Phillips, C.J.Skidmore, and I.O.Walker. 1978. The interaction of core histones with DNA: equilibrium binding studies. *Nucleic Acids Res.* **5**: 3643-3663.

Byvoet,P., G.R.Shepherd, J.M.Hardin, and B.J.Noland. 1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem. Biophys.* **148**: 558-567.

Cabral,A., P.Voskamp, A.M.Cleton-Jansen, A.South, D.Nizetic, and C.Backendorf. 2001. Structural organization and regulation of the small proline-rich family of cornified envelope precursors suggest a role in adaptive barrier function. *J. Biol. Chem.* **276**: 19231-19237.

Calogero,S., F.Grassi, A.Aguzzi, T.Voigtlander, P.Ferrier, S.Ferrari, and M.E.Bianchi. 1999. The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nat. Genet.* **22**: 276-280.

Candi,E., G.Melino, G.Mei, E.Tarcsa, S.I.Chung, L.N.Marekov, and P.M.Steinert. 1995. Biochemical, structural, and transglutaminase substrate properties of human loricrin, the major epidermal cornified cell envelope protein. *J. Biol. Chem.* **270**: 26382-26390.

Candi,E., R.Schmidt, and G.Melino. 2005. The cornified envelope: a model of cell death in the skin. *Nat. Rev. Mol. Cell Biol.* **6**: 328-340.

Candi,E., E.Tarcsa, W.W.Idler, T.Kartasova, L.N.Marekov, and P.M.Steinert. 1999. Transglutaminase cross-linking properties of the small proline-rich 1 family of cornified cell envelope proteins. Integration with loricrin. *J. Biol. Chem.* **274**: 7226-7237.

Cao,R., L.Wang, H.Wang, L.Xia, H.Erdjument-Bromage, P.Tempst, R.S.Jones, and Y.Zhang. 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**: 1039-1043.

Caplan,A., T.Kimura, H.Gould, and J.Allan. 1987. Perturbation of chromatin structure in the region of the adult beta-globin gene in chicken erythrocyte chromatin. *J. Mol. Biol.* **193**: 57-70.

Caron,H., S.B.van, M.M.van der, F.Baas, G.Riggins, S.P.van, M.C.Hermus, A.R.van, K.Boon, P.A.Voute, S.Heisterkamp, K.A.van, and R.Versteeg. 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**: 1289-1292.

Carson,S. and M.V.Wiles. 1993. Far upstream regions of class II MHC Ea are necessary for position-independent, copy-dependent expression of Ea transgene. *Nucleic Acids Res.* **21**: 2065-2072.

Carter,D., L.Chakalova, C.S.Osborne, Y.F.Dai, and P.Fraser. 2002. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* **32**: 623-626.

Castro,L.F., R.F.Furlong, and P.W.Holland. 2004. An antecedent of the MHC-linked genomic region in amphioxus. *Immunogenetics* **55**: 782-784.

Catez,F., D.T.Brown, T.Misteli, and M.Bustin. 2002. Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO Rep.* **3**: 760-766.

Catez,F., H.Yang, K.J.Tracey, R.Reeves, T.Misteli, and M.Bustin. 2004. Network of dynamic interactions between histone H1 and high-mobility-group proteins in chromatin. *Mol. Cell Biol.* **24**: 4321-4328.

Cereghini,S. and M.Yaniv. 1984. Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. *EMBO J.* **3**: 1243-1253.

Chadwick,B.P. and H.F.Willard. 2001. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J. Cell Biol.* **152**: 375-384.

Chadwick,B.P. and H.F.Willard. 2004. Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc. Natl. Acad. Sci. U. S. A* **101**: 17450-17455.

Chambeyron,S. and W.A.Bickmore. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**: 1119-1130.

Chambeyron,S., N.R.Da Silva, K.A.Lawson, and W.A.Bickmore. 2005. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**: 2215-2223.

Chow,C.M., A.Georgiou, H.Szutorisz, Maia e Silva, A.Pombo, I.Barahona, E.Dargelos, C.Canzonetta, and N.Dillon. 2005. Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep.* **6**: 354-360.

Christova,R., T.Jones, P.J.Wu, A.Bolzer, A.P.Costa-Pereira, D.Watling, I.M.Kerr, and D.Sheer. 2007. P-STAT1 mediates higher-order chromatin remodelling of the human MHC in response to IFNgamma. *J. Cell Sci.* **120**: 3262-3270.

Churchill,G.A. 2002. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32 Suppl**: 490-495.

Clark,D.J. and T.Kimura. 1990. Electrostatic mechanism of chromatin folding. *J. Mol. Biol.* **211**: 883-896.

Clark,K.L., E.D.Halay, E.Lai, and S.K.Burley. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**: 412-420.

Clarkson,M.J., J.R.Wells, F.Gibson, R.Saint, and D.J.Tremethick. 1999. Regions of variant histone His2AvD required for Drosophila development. *Nature* **399**: 694-697.

Clayton,A.L., S.Rose, M.J.Barratt, and L.C.Mahadevan. 2000. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *EMBO J.* **19**: 3714-3726.

Compton,J.G., J.J.DiGiovanna, K.A.Johnston, P.Fleckman, and S.J.Bale. 2002. Mapping of the associated phenotype of an absent granular layer in ichthyosis vulgaris to the epidermal differentiation complex on chromosome 1. *Exp. Dermatol.* **11**: 518-526.

Contzler,R., B.Favre, M.Huber, and D.Hohl. 2005. Cornulin, a new member of the "fused gene" family, is expressed during epidermal differentiation. *J. Invest Dermatol.* **124**: 990-997.

Cornforth,M.N., K.M.Greulich-Bode, B.D.Loucas, J.Arsuaga, M.Vazquez, R.K.Sachs, M.Bruckner, M.Molls, P.Hahnfeldt, L.Hlatky, and D.J.Brenner. 2002. Chromosomes are predominantly located randomly with respect to each other in interphase human cells. *J. Cell Biol.* **159**: 237-244.

Costanzi,C. and J.R.Pehrson. 1998. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393**: 599-601.

Craig,J.M. and W.A.Bickmore. 1994. The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.* **7**: 376-382.

Cremer,T. and C.Cremer. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**: 292-301.

Croft,J.A., J.M.Bridger, S.Boyle, P.Perry, P.Teague, and W.A.Bickmore. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**: 1119-1131.

Cuthbert,G.L., S.Daujat, A.W.Snowden, H.Erdjument-Bromage, T.Hagiwara, M.Yamada, R.Schneider, P.D.Gregory, P.Tempst, A.J.Bannister, and T.Kouzarides. 2004. Histone deimination antagonizes arginine methylation. *Cell* **118**: 545-553.

Czelusniak,J., M.Goodman, D.Hewett-Emmett, M.L.Weiss, P.J.Venta, and R.E.Tashian. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* **298**: 297-300.

Dai,X. and J.A.Segre. 2004. Transcriptional control of epidermal specification and differentiation. *Curr. Opin. Genet. Dev.* **14**: 485-491.

Dalal,Y., T.Furuyama, D.Vermaak, and S.Henikoff. 2007a. Structure, dynamics, and evolution of centromeric nucleosomes. *Proc. Natl. Acad. Sci. U. S. A.*

Dalal,Y., H.Wang, S.Lindsay, and S.Henikoff. 2007b. Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS. Biol.* **5**: e218.

Dalma-Weiszhausz,D.D., J.Warrington, E.Y.Tanimoto, and C.G.Miyada. 2006. The affymetrix GeneChip platform: an overview. *Methods Enzymol.* **410**: 3-28.

Danchin,E., V.Vitiello, A.Vienne, O.Richard, P.Gouret, M.F.McDermott, and P.Pontarotti. 2004. The major histocompatibility complex origin. *Immunol. Rev.* **198**: 216-232.

Das,C., K.Hizume, K.Batta, B.R.Kumar, S.S.Gadad, S.Ganguly, S.Lorain, A.Verreault, P.P.Sadhale, K.Takeyasu, and T.K.Kundu. 2006. Transcriptional coactivator PC4, a chromatin-associated protein, induces chromatin condensation. *Mol. Cell Biol.* **26**: 8303-8315.

Davey,C., S.Pennings, G.Meersseman, T.J.Wess, and J.Allan. 1995. Periodicity of strong nucleosome positioning sites around the chicken adult beta-globin gene may encode regularly spaced chromatin. *Proc. Natl. Acad. Sci. U. S. A* **92**: 11210-11214.

de Gobbi,M., E.Anguita, J.Hughes, J.A.Sloane-Stanley, J.A.Sharpe, C.M.Koch, I.Dunham, R.J.Gibbons, W.G.Wood, and D.R.Higgs. 2007. Tissue-specific histone modification and transcription factor binding in {alpha} globin gene expression. *Blood.*

de la Serna,I., Y.Ohkawa, and A.N.Imbalzano. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat. Rev. Genet.* **7**: 461-473.

De Leo,A.A., D.Wheeler, C.Lefevre, J.F.Cheng, R.Hope, J.Kuliwaba, K.R.Nicholas, M.Westerman, and J.A.Graves. 2005. Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial Sminthopsis macroura. *Cytogenet. Genome Res.* **108**: 333-341.

de Wit,E., F.Greil, and S.B.van. 2007. High-resolution mapping reveals links of HP1 with active and inactive chromatin components. *PLoS. Genet.* **3**: e38.

Dekker,J. 2007. GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol.* **8**: R116.

Dhalluin,C., J.E.Carlson, L.Zeng, C.He, A.K.Aggarwal, and M.M.Zhou. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**: 491-496.

Ding,H.F., M.Bustin, and U.Hansen. 1997. Alleviation of histone H1-mediated transcriptional repression and chromatin compaction by the acidic activation region in chromosomal protein HMG-14. *Mol. Cell Biol.* **17**: 5843-5855.

Dion,M.F., S.J.Altschuler, L.F.Wu, and O.J.Rando. 2005. Genomic characterization reveals a simple histone H4 acetylation code. *Proc. Natl. Acad. Sci. U. S. A* **102**: 5501-5506.

Djian,P., K.Easley, and H.Green. 2000. Targeted ablation of the murine involucrin gene. *J. Cell Biol.* **151**: 381-388.

Dominski,Z. and W.F.Marzluff. 1999. Formation of the 3' end of histone mRNA. *Gene* **239**: 1-14.

Dorigo,B., T.Schalch, K.Bystricky, and T.J.Richmond. 2003. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* **327**: 85-96.

Draves,P.H., P.T.Lowary, and J.Widom. 1992. Co-operative binding of the globular domain of histone H5 to DNA. *J. Mol. Biol.* **225**: 1105-1121.

Drissen,R., R.J.Palstra, N.Gillemans, E.Splinter, F.Grosveld, S.Philipsen, and L.W.de. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev.* **18**: 2485-2490.

Drouin,R., G.P.Holmquist, and C.L.Richer. 1994. High-resolution replication bands compared with morphologic G- and R-bands. *Adv. Hum. Genet.* **22**: 47-115.

Duboule,D. 2007. The rise and fall of Hox gene clusters. *Development* **134**: 2549-2560.

Duerre,J.A. and C.T.Lee. 1974. In vivo methylation and turnover of rat brain histones. *J. Neurochem.* **23**: 541-547.

Dunn,K.L. and J.R.Davie. 2005. Stimulation of the Ras-MAPK pathway leads to independent phosphorylation of histone H3 on serine 10 and 28. *Oncogene* **24**: 3492-3502.

Dyson,M.H., S.Thomson, M.Inagaki, H.Goto, S.J.Arthur, K.Nightingale, F.J.Iborra, and L.C.Mahadevan. 2005. MAP kinase-mediated phosphorylation of distinct pools

of histone H3 at S10 or S28 via mitogen- and stress-activated kinase 1/2. *J. Cell Sci.* **118**: 2247-2259.

El-Rifai,W., C.A.Moskaluk, M.K.Abdrabbo, J.Harper, C.Yoshida, G.J.Riggins, H.F.Frierson, Jr., and S.M.Powell. 2002. Gastric cancers overexpress S100A calcium-binding proteins. *Cancer Res.* **62**: 6823-6826.

Elder,J.T. and X.Zhao. 2002. Evidence for local control of gene expression in the epidermal differentiation complex. *Exp. Dermatol.* **11**: 406-412.

Faast,R., V.Thonglairoam, T.C.Schulz, J.Beall, J.R.Wells, H.Taylor, K.Matthaei, P.D.Rathjen, D.J.Tremethick, and I.Lyons. 2001. Histone variant H2A.Z is required for early mammalian development. *Curr. Biol.* **11**: 1183-1187.

Falciola,L., F.Spada, S.Calogero, G.Langst, R.Voit, I.Grummt, and M.E.Bianchi. 1997. High mobility group 1 protein is not stably associated with the chromosomes of somatic cells. *J. Cell Biol.* **137**: 19-26.

Fan,J.Y., F.Gordon, K.Luger, J.C.Hansen, and D.J.Tremethick. 2002. The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states. *Nat. Struct. Biol.* **9**: 172-176.

Fan,J.Y., D.Rangasamy, K.Luger, and D.J.Tremethick. 2004. H2A.Z alters the nucleosome surface to promote HP1alpha-mediated chromatin fiber folding. *Mol. Cell* **16**: 655-661.

Fan,Y., T.Nikitina, E.M.Morin-Kensicki, J.Zhao, T.R.Magnuson, C.L.Woodcock, and A.I.Skoultchi. 2003. H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol. Cell Biol.* **23**: 4559-4572.

Fan,Y., T.Nikitina, J.Zhao, T.J.Fleury, R.Bhattacharyya, E.E.Bouhassira, A.Stein, C.L.Woodcock, and A.I.Skoultchi. 2005. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**: 1199-1212.

Fan,Y., A.Sirotkin, R.G.Russell, J.Ayala, and A.I.Skoultchi. 2001. Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype. *Mol. Cell Biol.* **21**: 7933-7943.

Farris,S.D., E.D.Rubio, J.J.Moon, W.M.Gombert, B.H.Nelson, and A.Krumm. 2005. Transcription-induced chromatin remodeling at the c-myc gene involves the local exchange of histone H2A.Z. *J. Biol. Chem.* **280**: 25298-25303.

Ferreira,J., G.Paolella, C.Ramos, and A.I.Lamond. 1997. Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J. Cell Biol.* **139**: 1597-1610.

Ferrier,D.E. and C.Minguillon. 2003. Evolution of the Hox/ParaHox gene clusters. *Int. J. Dev. Biol.* **47**: 605-611.

Finch,J.T. and A.Klug. 1976. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A* **73**: 1897-1901.

Fisher,E.A. and G.Felsenfeld. 1986. Comparison of the folding of beta-globin and ovalbumin gene containing chromatin isolated from chicken oviduct and erythrocytes. *Biochemistry* **25**: 8010-8016.

Fletcher,T.M. and J.C.Hansen. 1995. Core histone tail domains mediate oligonucleosome folding and nucleosomal DNA organization through distinct molecular mechanisms. *J. Biol. Chem.* **270**: 25359-25362.

Forsberg,E.C., K.M.Downs, H.M.Christensen, H.Im, P.A.Nuzzi, and E.H.Bresnick. 2000. Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. *Proc. Natl. Acad. Sci. U. S. A* **97**: 14494-14499.

Francis,N.J., R.E.Kingston, and C.L.Woodcock. 2004. Chromatin compaction by a polycomb group protein complex. *Science* **306**: 1574-1577.

Franklin,S.G. and A.Zweidler. 1977. Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature* **266**: 273-275.

Fuchs,E. and S.Raghavan. 2002. Getting under the skin of epidermal morphogenesis. *Nat. Rev. Genet.* **3**: 199-209.

Garcia,B.A., S.B.Hake, R.L.Diaz, M.Kauer, S.A.Morris, J.Recht, J.Shabanowitz, N.Mishra, B.D.Strahl, C.D.Allis, and D.F.Hunt. 2007. Organismal differences in post-translational modifications in histones H3 and H4. *J. Biol. Chem.* **282**: 7641-7655.

Garcia-Ramirez,M., F.Dong, and J.Ausio. 1992. Role of the histone "tails" in the folding of oligonucleosomes depleted of histone H1. *J. Biol. Chem.* **267**: 19587-19595.

Garcia-Ramirez,M., C.Rocchini, and J.Ausio. 1995. Modulation of chromatin folding by histone acetylation. *J. Biol. Chem.* **270**: 17923-17928.

Gautier,T., D.W.Abbott, A.Molla, A.Verdel, J.Ausio, and S.Dimitrov. 2004. Histone variant H2ABbd confers lower stability to the nucleosome. *EMBO Rep.* **5**: 715-720.

Gazel,A., P.Ramphal, M.Rosdy, W.B.De, C.Tornier, N.Hosein, B.Lee, M.Tomic-Canic, and M.Blumenberg. 2003. Transcriptional profiling of epidermal keratinocytes: comparison of genes expressed in skin, cultured keratinocytes, and reconstituted epidermis, using large DNA microarrays. *J. Invest Dermatol.* **121**: 1459-1468.

Gerke,V. and K.Weber. 1985. The regulatory chain in the p36-kd substrate complex of viral tyrosine-specific protein kinases is related in sequence to the S-100 protein of glial cells. *EMBO J.* **4**: 2917-2920.

Ghirlando,R., M.D.Litt, M.N.Prioleau, F.Recillas-Targa, and G.Felsenfeld. 2004. Physical properties of a genomic condensed chromatin fragment. *J. Mol. Biol.* **336**: 597-605.

Gierman,H.J., M.H.Indemans, J.Koster, S.Goetze, J.Seppen, D.Geerts, D.R.van, and R.Versteeg. 2007. Domain-wide regulation of gene expression in the human genome. *Genome Res.* **17**: 1286-1295.

Gilbert,N. and J.Allan. 2001. Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U. S. A* **98**: 11949-11954.

Gilbert,N., S.Boyle, H.Fiegler, K.Woodfine, N.P.Carter, and W.A.Bickmore. 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**: 555-566.

Gilbert,N., S.Boyle, H.Sutherland, H.J.de Las, J.Allan, T.Jenuwein, and W.A.Bickmore. 2003. Formation of facultative heterochromatin in the absence of HP1. *EMBO J.* **22**: 5540-5550.

Gilbert,N., S.Gilchrist, and W.A.Bickmore. 2005. Chromatin organization in the mammalian nucleus. *Int. Rev. Cytol.* **242**: 283-336.

Gilbert,N., I.Thomson, S.Boyle, J.Allan, B.Ramsahoye, and W.A.Bickmore. 2007. DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J. Cell Biol.* **177**: 401-411.

Gilchrist,S., N.Gilbert, P.Perry, and W.A.Bickmore. 2004. Nuclear organization of centromeric domains is not perturbed by inhibition of histone deacetylases. *Chromosome. Res.* **12**: 505-516.

Gillemans,N., T.McMorrow, R.Tewari, A.W.Wai, C.Burgtorf, D.Drabek, N.Ventress, A.Langeveld, D.Higgs, K.Tan-Un, F.Grosveld, and S.Philipsen. 2003. Functional and comparative analysis of globin loci in pufferfish and humans. *Blood* **101**: 2842-2849.

Goetze,S., J.Mateos-Langerak, H.J.Gierman, L.W.de, O.Giromus, M.H.Indemans, J.Koster, V.Ondrej, R.Versteeg, and D.R.van. 2007. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol. Cell Biol.* **27**: 4475-4487.

Gollub,J. and G.Sherlock. 2006. Clustering microarray data. *Methods Enzymol.* **411**: 194-213.

Goodwin,G.H., C.Sanders, and E.W.Johns. 1973. A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. *Eur. J. Biochem.* **38**: 14-19.

Graham,F.L., J.Smiley, W.C.Russell, and R.Nairn. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**: 59-74.

Greil,F., W.E.de, H.J.Bussemaker, and S.B.van. 2007. HP1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. *EMBO J.* **26**: 741-751.

Grimbaldeston,M.A., C.L.Geczy, N.Tedla, J.J.Finlay-Jones, and P.H.Hart. 2003. S100A8 induction in keratinocytes by ultraviolet A irradiation is dependent on reactive oxygen intermediates. *J. Invest Dermatol.* **121**: 1168-1174.

Groft,C.M., S.N.Uljon, R.Wang, and M.H.Werner. 1998. Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly. *Proc. Natl. Acad. Sci. U. S. A* **95**: 9117-9122.

Hake,S.B. and C.D.Allis. 2006. Histone H3 variants and their potential role in indexing mammalian genomes: the "H3 barcode hypothesis". *Proc. Natl. Acad. Sci. U. S. A* **103**: 6428-6435.

Hake,S.B., B.A.Garcia, E.M.Duncan, M.Kauer, G.Dellaire, J.Shabanowitz, D.P.Bazett-Jones, C.D.Allis, and D.F.Hunt. 2006. Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J. Biol. Chem.* **281**: 559-568.

Hartman,P.G., G.E.Chapman, T.Moss, and E.M.Bradbury. 1977. Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule. *Eur. J. Biochem.* **77**: 45-51.

Hassig,C.A., T.C.Fleischer, A.N.Billin, S.L.Schreiber, and D.E.Ayer. 1997. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **89**: 341-347.

Hassig,C.A. and S.L.Schreiber. 1997. Nuclear histone acetylases and deacetylases and transcriptional regulation: HATs off to HDACs. *Curr. Opin. Chem. Biol.* **1**: 300-308.

Hazzalin,C.A. and L.C.Mahadevan. 2005. Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-fos and c-jun. *PLoS. Biol.* **3**: e393.

Hebbes,T.R., A.W.Thorne, and C.Crane-Robinson. 1988. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7**: 1395-1402.

Heighway,J., T.Knapp, L.Boyce, S.Brennand, J.K.Field, D.C.Betticher, D.Ratschiller, M.Gugger, M.Donovan, A.Lasek, and P.Rickert. 2002. Expression profiling of primary non-small cell lung cancer for target identification. *Oncogene* **21**: 7749-7763.

Heintzman,N.D., R.K.Stuart, G.Hon, Y.Fu, C.W.Ching, R.D.Hawkins, L.O.Barrera, C.S.Van, C.Qu, K.A.Ching, W.Wang, Z.Weng, R.D.Green, G.E.Crawford, and B.Ren. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**: 311-318.

Heitz,E. 1928. Das Heterochromatin der Moose. *Jahrb Wiss Botanik* **69**: 762-818.

Hennings,H., D.Michael, C.Cheng, P.Steinert, K.Holbrook, and S.H.Yuspa. 1980. Calcium regulation of growth and differentiation of mouse epidermal cells in culture. *Cell* **19**: 245-254.

Hewish,D.R. and L.A.Burgoyne. 1973. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* **52**: 504-510.

Higasa,K. and K.Hayashi. 2006. Periodicity of SNP distribution around transcription start sites. *BMC. Genomics* **7**: 66.

Hill,D.A. and R.Reeves. 1997. Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA. *Nucleic Acids Res.* **25**: 3523-3531.

Hoegg,S. and A.Meyer. 2005. Hox clusters as models for vertebrate genome evolution. *Trends Genet.* **21**: 421-424.

Horowitz,R.A., D.A.Agard, J.W.Sedat, and C.L.Woodcock. 1994. The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.* **125**: 1-10.

Horowitz-Scherer,R.A. and C.L.Woodcock. 2006. Organization of interphase chromatin. *Chromosoma* **115**: 1-14.

Horton,R., L.Wilming, V.Rand, R.C.Lovering, E.A.Bruford, V.K.Khodiyar, M.J.Lush, S.Povey, C.C.Talbot, Jr., M.W.Wright, H.M.Wain, J.Trowsdale, A.Ziegler, and S.Beck. 2004. Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**: 889-899.

Hsieh,H.L., B.W.Schafer, N.Sasaki, and C.W.Heizmann. 2003. Expression analysis of S100 proteins and RAGE in human tumors using tissue microarrays. *Biochem. Biophys. Res. Commun.* **307**: 375-381.

Huang,R.C. and J.Bonner. 1962. Histone, a suppressor of chromosomal RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A* **48**: 1216-1222.

Huang,S., M.Litt, and G.Felsenfeld. 2005. Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes Dev.* **19**: 1885-1893.

Hurst,L.D., C.Pal, and M.J.Lercher. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**: 299-310.

Ikuta,T., N.Yoshida, N.Satoh, and H.Saiga. 2004. Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc. Natl. Acad. Sci. U. S. A* **101**: 15118-15123.

Jackson,B., C.M.Tilli, M.J.Hardman, A.A.Avilion, M.C.MacLeod, G.S.Ashcroft, and C.Byrne. 2005. Late cornified envelope family in differentiating epithelia--response to calcium and ultraviolet irradiation. *J. Invest Dermatol.* **124**: 1062-1070.

Jackson,J.D. and M.A.Gorovsky. 2000. Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Res.* **28**: 3811-3816.

Jackson,V. 1978. Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. *Cell* **15**: 945-954.

James,T.C., J.C.Eissenberg, C.Craig, V.Dietrich, A.Hobson, and S.C.Elgin. 1989. Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. *Eur. J. Cell Biol.* **50**: 170-180.

Jang,S.I. and P.M.Steinert. 2002. Loricrin expression in cultured human keratinocytes is controlled by a complex interplay between transcription factors of the Sp1, CREB, AP1, and AP2 families. *J. Biol. Chem.* **277**: 42268-42279.

Janicki,S.M., T.Tsukamoto, S.E.Salghetti, W.P.Tansey, R.Sachidanandam, K.V.Prasanth, T.Ried, Y.Shav-Tal, E.Bertrand, R.H.Singer, and D.L.Spector. 2004. From silencing to gene expression: real-time analysis in single cells. *Cell* **116**: 683-698.

Jayaraman,L., N.C.Moorthy, K.G.Murthy, J.L.Manley, M.Bustin, and C.Prives. 1998. High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev.* **12**: 462-472.

Jeppesen,P. 1997. Histone acetylation: a possible mechanism for the inheritance of cell memory at mitosis. *Bioessays* **19**: 67-74.

Jin,C. and G.Felsenfeld. 2006. Distribution of histone H3.3 in hematopoietic cell lineages. *Proc. Natl. Acad. Sci. U. S. A* **103**: 574-579.

Jin,C. and G.Felsenfeld. 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**: 1519-1529.

Johnnidis,J.B., E.S.Venanzi, D.J.Taxman, J.P.Ting, C.O.Benoist, and D.J.Mathis. 2005. Chromosomal clustering of genes controlled by the aire transcription factor. *Proc. Natl. Acad. Sci. U. S. A* **102**: 7233-7238.

Johnston,R., B.Wang, R.Nuttall, M.Doctolero, P.Edwards, J.Lu, M.Vainer, H.Yue, X.Wang, J.Minor, C.Chan, A.Lash, T.Goralski, M.Parisi, B.Oliver, and S.Eastman. 2004. FlyGEM, a full transcriptome array platform for the Drosophila community. *Genome Biol.* **5**: R19.

Kapranov,P., S.E.Cawley, J.Drenkow, S.Bekiranov, R.L.Strausberg, S.P.Fodor, and T.R.Gingeras. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916-919.

Kendall,M. 1938. A New Measure of Rank Correlation. *Biometrika* **30**: 81-89.

Kim,T.H., Z.K.Abdullaev, A.D.Smith, K.A.Ching, D.I.Loukinov, R.D.Green, M.Q.Zhang, V.V.Lobanenkov, and B.Ren. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**: 1231-1245.

Kim,T.H. and B.Ren. 2006. Genome-Wide Analysis of Protein-DNA Interactions. *Annu. Rev. Genomics Hum. Genet.* **7**: 81-102.

Kimura,T., F.C.Mills, J.Allan, and H.Gould. 1983. Selective unfolding of erythroid chromatin in the region of the active beta-globin gene. *Nature* **306**: 709-712.

Klose,R.J. and Y.Zhang. 2007. Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* **8**: 307-318.

Kmita,M. and D.Duboule. 2003. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**: 331-333.

Knezetic,J.A. and D.S.Luse. 1986. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell* **45**: 95-104.

Koch,C.M., R.M.Andrews, P.Flicek, S.C.Dillon, U.Karaoz, G.K.Clelland, S.Wilcox, D.M.Beare, J.C.Fowler, P.Couttet, K.D.James, G.C.Lefebvre, A.W.Bruce, O.M.Dovey, P.D.Ellis, P.Dhami, C.F.Langford, Z.Weng, E.Birney, N.P.Carter, D.Vetrie, and I.Dunham. 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.* **17**: 691-707.

Koch,P.J., P.A.de Viragh, E.Scharer, D.Bundman, M.A.Longley, J.Bickenbach, Y.Kawachi, Y.Suga, Z.Zhou, M.Huber, D.Hohl, T.Kartasova, M.Jarnik, A.C.Steven, and D.R.Roop. 2000. Lessons from loricrin-deficient mice: compensatory mechanisms maintaining skin barrier function in the absence of a major cornified envelope protein. *J. Cell Biol.* **151**: 389-400.

Koizumi,H., T.Kartasova, H.Tanaka, A.Ohkawara, and T.Kuroki. 1996. Differentiation-associated localization of small proline-rich protein in normal and diseased human skin. *Br. J. Dermatol.* **134**: 686-692.

Kouzarides,T. 2007. Chromatin modifications and their function. *Cell* **128**: 693-705.

Krawczyk,M., N.Peyraud, N.Rybtsova, K.Masternak, P.Bucher, E.Barras, and W.Reith. 2004. Long distance control of MHC class II expression by multiple distal enhancers regulated by regulatory factor X complex and CIITA. *J. Immunol.* **173**: 6200-6210.

Krieg,P., M.Schuppler, R.Koesters, A.Mincheva, P.Lichter, and F.Marks. 1997. Repetin (Rptn), a new member of the "fused gene" subgroup within the S100 gene family encoding a murine epidermal differentiation protein. *Genomics* **43**: 339-348.

Kumanovics,A., T.Takada, and K.F.Lindahl. 2003. Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.* **21**: 629-657.

Kuo,M.H., J.Zhou, P.Jambeck, M.E.Churchill, and C.D.Allis. 1998. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev.* **12**: 627-639.

Kuzmichev,A., K.Nishioka, H.Erdjument-Bromage, P.Tempst, and D.Reinberg. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**: 2893-2905.

Lachner,M., D.O'Carroll, S.Rea, K.Mechtler, and T.Jenuwein. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116-120.

Lander,E.S., L.M.Linton, B.Birren, C.Nusbaum, M.C.Zody, J.Baldwin, K.Devon, K.Dewar, M.Doyle, W.FitzHugh, R.Funke, D.Gage, K.Harris, A.Heaford, J.Howland, L.Kann, J.Lehoczky, R.LeVine, P.McEwan, K.McKernan, J.Meldrim, J.P.Mesirov, C.Miranda, W.Morris, J.Naylor, C.Raymond, M.Rosetti, R.Santos, A.Sheridan, C.Sougnez, N.Stange-Thomann, N.Stojanovic, A.Subramanian, D.Wyman, J.Rogers, J.Sulston, R.Ainscough, S.Beck, D.Bentley, J.Burton, C.Clee, N.Carter, A.Coulson, R.Deadman, P.Deloukas, A.Dunham, I.Dunham, R.Durbin, L.French, D.Grafham, S.Gregory, T.Hubbard, S.Humphray, A.Hunt, M.Jones, C.Lloyd, A.McMurray, L.Matthews, S.Mercer, S.Milne, J.C.Mullikin, A.Mungall, R.Plumb, M.Ross, R.Shownkeen, S.Sims, R.H.Waterston, R.K.Wilson, L.W.Hillier, J.D.McPherson, M.A.Marra, E.R.Mardis, L.A.Fulton, A.T.Chinwalla, K.H.Pepin, W.R.Gish, S.L.Chissoe, M.C.Wendl, K.D.Delehaunty, T.L.Miner, A.Delehaunty, J.B.Kramer, L.L.Cook, R.S.Fulton, D.L.Johnson, P.J.Minx, S.W.Clifton, T.Hawkins, E.Branscomb, P.Predki, P.Richardson, S.Wenning, T.Slezak, N.Doggett, J.F.Cheng, A.Olsen, S.Lucas, C.Elkin, E.Uberbacher, M.Frazier, R.A.Gibbs, D.M.Muzny, S.E.Scherer, J.B.Bouck, E.J.Sodergren, K.C.Worley, C.M.Rives, J.H.Gorrell, M.L.Metzker, S.L.Naylor, R.S.Kucherlapati, D.L.Nelson, G.M.Weinstock, Y.Sakaki, A.Fujiyama, M.Hattori, T.Yada, A.Toyoda, T.Itoh, C.Kawagoe, H.Watanabe, Y.Totoki, T.Taylor, J.Weissenbach, R.Heilig, W.Saurin, F.Artiguenave, P.Brottier, T.Bruls, E.Pelletier, C.Robert, P.Wincker, D.R.Smith, L.Doucette-Stamm, M.Rubenfield, K.Weinstock, H.M.Lee, J.Dubois, A.Rosenthal, M.Platzer, G.Nyakatura, S.Taudien, A.Rump, H.Yang, J.Yu, J.Wang, G.Huang, J.Gu, L.Hood, L.Rowen, A.Madan, S.Qin, R.W.Davis, N.A.Federspiel, A.P.Abola, M.J.Proctor, R.M.Myers, J.Schmutz, M.Dickson, J.Grimwood, D.R.Cox, M.V.Olson, R.Kaul, C.Raymond, N.Shimizu, K.Kawasaki, S.Minoshima, G.A.Evans, M.Athanasiou, R.Schultz, B.A.Roe, F.Chen, H.Pan, J.Ramser, H.Lehrach, R.Reinhardt, W.R.McCombie, B.M.de la, N.Dedhia, H.Blocker, K.Hornischer, G.Nordsiek, R.Agarwala, L.Aravind, J.A.Bailey, A.Bateman, S.Batzoglou, E.Birney, P.Bork, D.G.Brown, C.B.Burge, L.Cerutti, H.C.Chen, D.Church, M.Clamp, R.R.Copley, T.Doerks, S.R.Eddy, E.E.Eichler, T.S.Furey, J.Galagan, J.G.Gilbert, C.Harmon, Y.Hayashizaki, D.Haussler, H.Hermjakob, K.Hokamp, W.Jang, L.S.Johnson, T.A.Jones, S.Kasif, A.Kaspryzk, S.Kennedy, W.J.Kent, P.Kitts, E.V.Koonin, I.Korf, D.Kulp, D.Lancet, T.M.Lowe, A.McLysaght, T.Mikkelsen, J.V.Moran, N.Mulder, V.J.Pollara, C.P.Ponting, G.Schuler, J.Schultz, G.Slater, A.F.Smit, E.Stupka, J.Szustakowski, D.Thierry-Mieg, J.Thierry-Mieg, L.Wagner, J.Wallis, R.Wheeler, A.Williams, Y.I.Wolf, K.H.Wolfe, S.P.Yang, R.F.Yeh, F.Collins, M.S.Guyer, J.Peterson, A.Felsenfeld, K.A.Wetterstrand, A.Patrinos, M.J.Morgan, J.P.de,

J.J.Catanese, K.Osoegawa, H.Shizuya, S.Choi, and Y.J.Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Lapi,E., A.Iovino, G.Fontemaggi, A.R.Soliera, S.Iacovelli, A.Sacchi, G.Rechavi, D.Givol, G.Blandino, and S.Strano. 2006. S100A2 gene is a direct transcriptional target of p53 homologues during keratinocyte differentiation. *Oncogene* **25**: 3628-3637.

Lechler,T. and E.Fuchs. 2005. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature* **437**: 275-280.

Lee,H., R.Habas, and C.bate-Shen. 2004. MSX1 cooperates with histone H1b for inhibition of transcription and myogenesis. *Science* **304**: 1675-1678.

Lennox,R.W. and L.H.Cohen. 1983. The histone H1 complements of dividing and nondividing cells of the mouse. *J. Biol. Chem.* **258**: 262-268.

Lercher,M.J., A.O.Urrutia, and L.D.Hurst. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31**: 180-183.

Lesniak,W., L.P.Slomnicki, and J.Kuznicki. 2007. Epigenetic Control of the S100A6 (Calcyclin) Gene Expression. *J. Invest Dermatol.*

Leuba,S.H., G.Yang, C.Robert, B.Samori, H.K.van, J.Zlatanova, and C.Bustamante. 1994. Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. *Proc. Natl. Acad. Sci. U. S. A* **91**: 11621-11625.

Li,B., S.G.Pattenden, D.Lee, J.Gutierrez, J.Chen, C.Seidel, J.Gerton, and J.L.Workman. 2005. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl. Acad. Sci. U. S. A* **102**: 18385-18390.

Li,Y., J.R.Danzer, P.Alvarez, A.S.Belmont, and L.L.Wallrath. 2003. Effects of tethering HP1 to euchromatic regions of the Drosophila genome. *Development* **130**: 1817-1824.

Li,Y., D.A.Kirschmann, and L.L.Wallrath. 2002. Does heterochromatin protein 1 always follow code? *Proc. Natl. Acad. Sci. U. S. A* **99 Suppl 4**: 16462-16469.

Liang,G., J.C.Lin, V.Wei, C.Yoo, J.C.Cheng, C.T.Nguyen, D.J.Weisenberger, G.Egger, D.Takai, F.A.Gonzales, and P.A.Jones. 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. U. S. A* **101**: 7357-7362.

Lio,P. 2003. Wavelets in bioinformatics and computational biology: state of art and perspectives. *Bioinformatics.* **19**: 2-9.

Litt,M.D., M.Simpson, M.Gaszner, C.D.Allis, and G.Felsenfeld. 2001. Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* **293**: 2453-2455.

Logsdon,C.D., D.M.Simeone, C.Binkley, T.Arumugam, J.K.Greenson, T.J.Giordano, D.E.Misek, R.Kuick, and S.Hanash. 2003. Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer. *Cancer Res.* **63**: 2649-2657.

Lorch,Y., J.W.LaPointe, and R.D.Kornberg. 1987. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**: 203-210.

Lowary,P.T. and J.Widom. 1997. Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl. Acad. Sci. U. S. A* **94**: 1183-1188.

Luger,K., A.W.Mader, R.K.Richmond, D.F.Sargent, and T.J.Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**: 251-260.

Mahy,N.L., P.E.Perry, and W.A.Bickmore. 2002a. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J. Cell Biol.* **159**: 753-763.

Mahy,N.L., P.E.Perry, S.Gilchrist, R.A.Baldock, and W.A.Bickmore. 2002b. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J. Cell Biol.* **157**: 579-589.

Mainguy,G., J.Koster, J.Woltering, H.Jansen, and A.Durston. 2007. Extensive polycistronism and antisense transcription in the Mammalian Hox clusters. *PLoS. ONE.* **2**: e356.

Mak,W., J.Baxter, J.Silva, A.E.Newall, A.P.Otte, and N.Brockdorff. 2002. Mitotically stable association of polycomb group proteins eed and enx1 with the inactive x chromosome in trophoblast stem cells. *Curr. Biol.* **12**: 1016-1020.

Makino,T., M.Takaishi, M.Morohashi, and N.H.Huh. 2001. Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin. *J. Biol. Chem.* **276**: 47445-47452.

Malik,H.S. and S.Henikoff. 2003. Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10**: 882-891.

Manuelidis,L. 1985. Individual interphase chromosome domains revealed by in situ hybridization. *Hum. Genet.* **71**: 288-293.

Marenholz,I., C.W.Heizmann, and G.Fritz. 2004. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem. Biophys. Res. Commun.* **322**: 1111-1122.

Marshall,D., M.J.Hardman, K.M.Nield, and C.Byrne. 2001. Differentially expressed late constituents of the epidermal cornified envelope. *Proc. Natl. Acad. Sci. U. S. A* **98**: 13031-13036.

Martens,J.H., R.J.O'Sullivan, U.Braunschweig, S.Opravil, M.Radolf, P.Steinlein, and T.Jenuwein. 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**: 800-812.

Martin,N., S.Patel, and J.A.Segre. 2004. Long-range comparison of human and mouse Sprr loci to identify conserved noncoding sequences involved in coordinate regulation. *Genome Res.* **14**: 2430-2438.

Marzluff,W.F., P.Gongidi, K.R.Woods, J.Jin, and L.J.Maltais. 2002. The human and mouse replication-dependent histone genes. *Genomics* **80**: 487-498.

Mason,M.M., J.A.Grasso, O.Gavrilova, and M.Reitman. 1996. Identification of functional elements of the chicken epsilon-globin promoter involved in stage-specific interaction with the beta/epsilon enhancer. *J. Biol. Chem.* **271**: 25459-25467.

Mason,M.M., E.Lee, H.Westphal, and M.Reitman. 1995. Expression of the chicken beta-globin gene cluster in mice: correct developmental expression and distributed control. *Mol. Cell Biol.* **15**: 407-414.

Masternak,K., N.Peyraud, M.Krawczyk, E.Barras, and W.Reith. 2003. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat. Immunol.* **4**: 132-137.

Mathis,D. and C.Benoist. 2007. A decade of AIRE. *Nat. Rev. Immunol.* **7**: 645-650.

Mathur,P., B.Murray, T.Crowell, H.Gardner, N.Allaire, Y.M.Hsu, G.Thill, and J.P.Carulli. 2004. Murine peptidoglycan recognition proteins PglyrpIalpha and PglyrpIbeta are encoded in the epidermal differentiation complex and are expressed in epidermal and hematopoietic tissues. *Genomics* **83**: 1151-1163.

Matsuki,M., F.Yamashita, A.Ishida-Yamamoto, K.Yamada, C.Kinoshita, S.Fushiki, E.Ueda, Y.Morishima, K.Tabata, H.Yasuno, M.Hashida, H.Iizuka, M.Ikawa, M.Okabe, G.Kondoh, T.Kinoshita, J.Takeda, and K.Yamanishi. 1998. Defective stratum corneum and early neonatal death in mice lacking the gene for transglutaminase 1 (keratinocyte transglutaminase). *Proc. Natl. Acad. Sci. U. S. A* **95**: 1044-1049.

McGhee,J.D. and G.Felsenfeld. 1980. Nucleosome structure. *Annu. Rev. Biochem.* **49**: 1115-1156.

McKay,D.B. and T.A.Steitz. 1981. Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA. *Nature* **290**: 744-749.

Meersseman,G., S.Pennings, and E.M.Bradbury. 1992. Mobile nucleosomes--a general behavior. *EMBO J.* **11**: 2951-2959.

Mehul,B., D.Asselineau, D.Bernard, J.Leclaire, M.Regnier, R.Schmidt, and F.Bernerd. 2004. Gene expression profiles of three different models of reconstructed human epidermis and classical cultures of keratinocytes using cDNA arrays. *Arch. Dermatol. Res.* **296**: 145-156.

Mello,J.A. and G.Almouzni. 2001. The ins and outs of nucleosome assembly. *Curr. Opin. Genet. Dev.* **11**: 136-141.

Meneghini,M.D., M.Wu, and H.D.Madhani. 2003. Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112**: 725-736.

Menon,G.K., P.M.Elias, S.H.Lee, and K.R.Feingold. 1992. Localization of calcium in murine epidermis following disruption and repair of the permeability barrier. *Cell Tissue Res.* **270**: 503-512.

Metivier,R., G.Penot, M.R.Hubner, G.Reid, H.Brand, M.Kos, and F.Gannon. 2003. Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**: 751-763.

Metzger,E., M.Wissmann, N.Yin, J.M.Muller, R.Schneider, A.H.Peters, T.Gunther, R.Buettner, and R.Schule. 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**: 436-439.

Mirkovitch,J., M.E.Mirault, and U.K.Laemmli. 1984. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39**: 223-232.

Mischke,D. 1998. The complexity of gene families involved in epithelial differentiation. Keratin genes and the epidermal differentiation complex. *Subcell. Biochem.* **31**: 71-104.

Mischke,D., B.P.Korge, I.Marenholz, A.Volz, and A.Ziegler. 1996. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J. Invest Dermatol.* **106**: 989-992.

Misteli,T., A.Gunjan, R.Hock, M.Bustin, and D.T.Brown. 2000. Dynamic binding of histone H1 to chromatin in living cells. *Nature* **408**: 877-881.

Mito,Y., J.G.Henikoff, and S.Henikoff. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**: 1090-1097.

Morey,C., N.R.Da Silva, P.Perry, and W.A.Bickmore. 2007. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134**: 909-919.

Morgan,R. 2002. A common element in epidermal expression? *Trends Genet.* **18**: 184-185.

Morgan,R., M.H.Hooiveld, R.P.In der, and A.J.Durston. 1999. A conserved 30 base pair element in the Wnt-5a promoter is sufficient both to drive its' early embryonic expression and to mediate its' repression by otx2. *Mech. Dev.* **85**: 97-102.

Morley,S.M., M.D'Alessandro, C.Sexton, E.L.Rugg, H.Navsaria, C.S.Shemanko, M.Huber, D.Hohl, A.I.Heagerty, I.M.Leigh, and E.B.Lane. 2003. Generation and characterization of epidermolysis bullosa simplex cell lines: scratch assays show faster migration with disruptive keratin mutations. *Br. J. Dermatol.* **149**: 46-58.

Nagy,P.L., M.L.Cleary, P.O.Brown, and J.D.Lieb. 2003. Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc. Natl. Acad. Sci. U. S. A* **100**: 6364-6369.

Neal,S.J., M.L.Gibson, A.K.So, and J.T.Westwood. 2003. Construction of a cDNA-based microarray for Drosophila melanogaster: a comparison of gene transcription profiles from SL2 and Kc167 cells. *Genome* **46**: 879-892.

Neal,S.J. and J.T.Westwood. 2006. Optimizing experiment and analysis parameters for spotted microarrays. *Methods Enzymol.* **410**: 203-221.

Nielsen,A.L., J.A.Ortiz, J.You, M.Oulad-Abdelghani, R.Khechumian, A.Gansmuller, P.Chambon, and R.Losson. 1999. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J.* **18**: 6385-6395.

Nomura,T., A.Sandilands, M.Akiyama, H.Liao, A.T.Evans, K.Sakai, M.Ota, H.Sugiura, K.Yamamoto, H.Sato, C.N.Palmer, F.J.Smith, W.H.McLean, and H.Shimizu. 2007. Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J. Allergy Clin. Immunol.* **119**: 434-440.

O'Neill,L.P. and B.M.Turner. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods* **31**: 76-82.

Ogryzko,V.V., R.L.Schiltz, V.Russanova, B.H.Howard, and Y.Nakatani. 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**: 953-959.

Oliva,R., D.P.Bazett-Jones, L.Locklear, and G.H.Dixon. 1990. Histone hyperacetylation can induce unfolding of the nucleosome core particle. *Nucleic Acids Res.* **18**: 2739-2747.

Orlando,V., H.Strutt, and R.Paro. 1997. Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **11**: 205-214.

Osoegawa,K., A.G.Mammoser, C.Wu, E.Frengen, C.Zeng, J.J.Catanese, and P.J.de Jong. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* **11**: 483-496.

Oudet,P., M.Gross-Bellard, and P.Chambon. 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**: 281-300.

Palmer,C.N., A.D.Irvine, A.Terron-Kwiatkowski, Y.Zhao, H.Liao, S.P.Lee, D.R.Goudie, A.Sandilands, L.E.Campbell, F.J.Smith, G.M.O'Regan, R.M.Watson, J.E.Cecil, S.J.Bale, J.G.Compton, J.J.DiGiovanna, P.Fleckman, S.Lewis-Jones, G.Arseculeratne, A.Sergeant, C.S.Munro, H.B.El, K.McElreavey, L.B.Halkjaer, H.Bisgaard, S.Mukhopadhyay, and W.H.McLean. 2006. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat. Genet.* **38**: 441-446.

Palmer,D.K., K.O'Day, H.L.Trong, H.Charbonneau, and R.L.Margolis. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc. Natl. Acad. Sci. U. S. A* **88**: 3734-3738.

Palstra,R.J., B.Tolhuis, E.Splinter, R.Nijmeijer, F.Grosveld, and L.W.de. 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**: 190-194.

Paramio,J.M., S.Lain, C.Segrelles, E.B.Lane, and J.L.Jorcano. 1998. Differential expression and functionally co-operative roles for the retinoblastoma family of proteins in epidermal differentiation. *Oncogene* **17**: 949-957.

Park,Y.J., P.N.Dyer, D.J.Tremethick, and K.Luger. 2004. A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. *J. Biol. Chem.* **279**: 24274-24282.

Patel,N.H. 2004. Evolutionary biology: time, space and genomes. *Nature* **431**: 28-29.

Patel,S., T.Kartasova, and J.A.Segre. 2003. Mouse Sprr locus: a tandem array of coordinately regulated genes. *Mamm. Genome* **14**: 140-148.

Patrone,G., F.Puppo, R.Cusano, M.Scaranari, I.Ceccherini, A.Puliti, and R.Ravazzolo. 2000. Nuclear run-on assay using biotin labeling, magnetic bead capture and analysis by fluorescence-based RT-PCR. *Biotechniques* **29**: 1012-1017.

Peckham,H.E., R.E.Thurman, Y.Fu, J.A.Stamatoyannopoulos, W.S.Noble, K.Struhl, and Z.Weng. 2007. Nucleosome positioning signals in genomic DNA. *Genome Res.* **17**: 1170-1177.

Perez-Burgos,L., A.H.Peters, S.Opravil, M.Kauer, K.Mechtler, and T.Jenuwein. 2004. Generation and characterization of methyl-lysine histone antibodies. *Methods Enzymol.* **376**: 234-254.

Peters,A.H., S.Kubicek, K.Mechtler, R.J.O'Sullivan, A.A.Derijck, L.Perez-Burgos, A.Kohlmaier, S.Opravil, M.Tachibana, Y.Shinkai, J.H.Martens, and T.Jenuwein. 2003. Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**: 1577-1589.

Phair,R.D. and T.Misteli. 2000. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**: 604-609.

Phimister,B. 1999. The Chipping Forecast. pp. 1-60.

Pinkel,D. and D.G.Albertson. 2005. Array comparative genomic hybridization and its applications in cancer. *Nat. Genet.* **37 Suppl**: S11-S17.

Plant,K.E., S.J.Routledge, and N.J.Proudfoot. 2001. Intergenic transcription in the human beta-globin gene cluster. *Mol. Cell Biol.* **21**: 6507-6514.

Presland,R.B., M.Tomic-Canic, S.P.Lewis, and B.A.Dale. 2001. Regulation of human profilaggrin promoter activity in cultured epithelial cells by retinoic acid and glucocorticoids. *J. Dermatol. Sci.* **27**: 192-205.

Price,T.S., R.Regan, R.Mott, A.Hedman, B.Honey, R.J.Daniels, L.Smith, A.Greenfield, A.Tiganescu, V.Buckle, N.Ventress, H.Ayyub, A.Salhan, S.Pedraza-Diaz, J.Broxholme, J.Ragoussis, D.R.Higgs, J.Flint, and S.J.Knight. 2005. SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res.* **33**: 3455-3464.

Prunell,A. 1982. Nucleosome reconstitution on plasmid-inserted poly(dA) . poly(dT). *EMBO J.* **1**: 173-179.

Quackenbush,J. 2002. Microarray data normalization and transformation. *Nat. Genet.* **32 Suppl**: 496-501.

Radic,M.Z., M.Saghbini, T.S.Elton, R.Reeves, and B.A.Hamkalo. 1992. Hoechst 33258, distamycin A, and high mobility group protein I (HMG-I) compete for binding to mouse satellite DNA. *Chromosoma* **101**: 602-608.

Radoja,N., A.Gazel, T.Banno, S.Yano, and M.Blumenberg. 2006. Transcriptional profiling of epidermal differentiation. *Physiol Genomics* **27**: 65-78.

Ramakrishnan,V., J.T.Finch, V.Graziano, P.L.Lee, and R.M.Sweet. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**: 219-223.

Rando,O.J. and K.Ahmad. 2007. Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.* **19**: 250-256.

Rasheed,S., W.A.Nelson-Rees, E.M.Toth, P.Arnstein, and M.B.Gardner. 1974. Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* **33**: 1027-1033.

Raval,A., T.K.Howcroft, J.D.Weissman, S.Kirshner, X.S.Zhu, K.Yokoyama, J.Ting, and D.S.Singer. 2001. Transcriptional coactivator, CIITA, is an acetyltransferase that bypasses a promoter requirement for TAF(II)250. *Mol. Cell* **7**: 105-115.

Rea,S., F.Eisenhaber, D.O'Carroll, B.D.Strahl, Z.W.Sun, M.Schmid, S.Opravil, K.Mechtler, C.P.Ponting, C.D.Allis, and T.Jenuwein. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593-599.

Rees,J. 2007. The Genetics of Common Skin Diseases. in *Genes and Common Diseases* (ed. A.Wright and N.Hastie), pp. 427-438. Cambridge University Press.

Remboutsika,E., Y.Lutz, A.Gansmuller, J.L.Vonesch, R.Losson, and P.Chambon. 1999. The putative nuclear receptor mediator TIF1alpha is tightly associated with euchromatin. *J. Cell Sci.* **112 (Pt 11)**: 1671-1683.

Rill,R. and K.E.Van Holde. 1973. Properties of nuclease-resistant fragments of calf thymus chromatin. *J. Biol. Chem.* **248**: 1080-1083.

Rinn,J.L., M.Kertesz, J.K.Wang, S.L.Squazzo, X.Xu, S.A.Brugmann, L.H.Goodnough, J.A.Helms, P.J.Farnham, E.Segal, and H.Y.Chang. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311-1323.

Robinson,N.A., S.Lapic, J.F.Welter, and R.L.Eckert. 1997. S100A11, S100A10, annexin I, desmosomal proteins, small proline-rich proteins, plasminogen activator inhibitor-2, and involucrin are components of the cornified envelope of cultured human epidermal keratinocytes. *J. Biol. Chem.* **272**: 12035-12046.

Robinson,P.J. and D.Rhodes. 2006. Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr. Opin. Struct. Biol.* **16**: 336-343.

Roh,T.Y., S.Cuddapah, K.Cui, and K.Zhao. 2006. The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U. S. A* **103**: 15782-15787.

Roh,T.Y., S.Cuddapah, and K.Zhao. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**: 542-552.

Roh,T.Y., W.C.Ngau, K.Cui, D.Landsman, and K.Zhao. 2004. High-resolution genome-wide mapping of histone modifications. *Nat. Biotechnol.* **22**: 1013-1016.

Rosenberger,S., I.S.Thorey, S.Werner, and P.Boukamp. 2007. A novel regulator of telomerase. S100A8 mediates differentiation-dependent and calcium-induced inhibition of telomerase activity in the human epidermal keratinocyte line HaCaT. *J. Biol. Chem.* **282**: 6126-6135.

Ruthenburg,A.J., C.D.Allis, and J.Wysocka. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**: 15-30.

Ryckman,C., K.Vandal, P.Rouleau, M.Talbot, and P.A.Tessier. 2003. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *J. Immunol.* **170**: 3233-3242.

Sachs,R.K., E.G.van den, B.Trask, H.Yokota, and J.E.Hearst. 1995. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci. U. S. A* **92**: 2710-2714.

Sadoni,N., S.Langer, C.Fauth, G.Bernardi, T.Cremer, B.M.Turner, and D.Zink. 1999. Nuclear organization of mammalian genomes. Polar chromosome territories build up functionally distinct higher order compartments. *J. Cell Biol.* **146**: 1211-1226.

Sadoni,N. and D.Zink. 2004. Nascent RNA synthesis in the context of chromatin architecture. *Chromosome. Res.* **12**: 439-451.

Sahasrabuddhe,C.G. and K.E.Van Holde. 1974. The effect of trypsin on nuclease-resistant chromatin fragments. *J. Biol. Chem.* **249**: 152-156.

Sakaguchi,M., M.Miyazaki, M.Takaishi, Y.Sakaguchi, E.Makino, N.Kataoka, H.Yamada, M.Namba, and N.H.Huh. 2003. S100C/A11 is a key mediator of Ca(2+)-induced growth inhibition of human epidermal keratinocytes. *J. Cell Biol.* **163**: 825-835.

Sambrook,J. and D.W.Russell. *Molecular Cloning: A Laboratory Manual (3rd Edition)*.

Sandilands,A., A.Terron-Kwiatkowski, P.R.Hull, G.M.O'Regan, T.H.Clayton, R.M.Watson, T.Carrick, A.T.Evans, H.Liao, Y.Zhao, L.E.Campbell, M.Schmuth, R.Gruber, A.R.Janecke, P.M.Elias, M.A.van Steensel, I.Nagtzaam, G.M.van, P.M.Steijlen, C.S.Munro, D.G.Bradley, C.N.Palmer, F.J.Smith, W.H.McLean, and A.D.Irvine. 2007. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat. Genet.* **39**: 650-654.

Sandman,K., S.L.Pereira, and J.N.Reeve. 1998. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell Mol. Life Sci.* **54**: 1350-1364.

Sandoval,J., J.L.Rodriguez, G.Tur, G.Serviddio, J.Pereda, A.Boukaba, J.Sastre, L.Torres, L.Franco, and G.Lopez-Rodas. 2004. RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nucleic Acids Res.* **32**: e88.

Santos-Rosa,H., R.Schneider, A.J.Bannister, J.Sherriff, B.E.Bernstein, N.C.Emre, S.L.Schreiber, J.Mellor, and T.Kouzarides. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature* **419**: 407-411.

Saris,C.J., T.Kristensen, P.D'Eustachio, L.J.Hicks, D.J.Noonan, T.Hunter, and B.F.Tack. 1987. cDNA sequence and tissue distribution of the mRNA for bovine and murine p11, the S100-related light chain of the protein-tyrosine kinase substrate p36 (calpactin I). *J. Biol. Chem.* **262**: 10663-10671.

Schalch,T., S.Duda, D.F.Sargent, and T.J.Richmond. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**: 138-141.

Schneider,R., A.J.Bannister, F.A.Myers, A.W.Thorne, C.Crane-Robinson, and T.Kouzarides. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell Biol.* **6**: 73-77.

Schroth,G.P., P.Yau, B.S.Imai, J.M.Gatewood, and E.M.Bradbury. 1990. A NMR study of mobility in the histone octamer. *FEBS Lett.* **268**: 117-120.

Schubeler,D., D.M.MacAlpine, D.Scalzo, C.Wirbelauer, C.Kooperberg, L.F.van, D.E.Gottschling, L.P.O'Neill, B.M.Turner, J.Delrow, S.P.Bell, and M.Groudine. 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* **18**: 1263-1271.

Schwarz,P.M., A.Felthauser, T.M.Fletcher, and J.C.Hansen. 1996. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry* **35**: 4009-4015.

Seale,R.L., A.T.Annunziato, and R.D.Smith. 1983. High mobility group proteins: abundance, turnover, and relationship to transcriptionally active chromatin. *Biochemistry* **22**: 5008-5015.

Segal,E., Y.Fondufe-Mittendorf, L.Chen, A.Thastrom, Y.Field, I.K.Moore, J.P.Wang, and J.Widom. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772-778.

Sekinger,E.A., Z.Moqtaderi, and K.Struhl. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**: 735-748.

Seo,H.C., R.B.Edvardsen, A.D.Maeland, M.Bjordal, M.F.Jensen, A.Hansen, M.Flaat, J.Weissenbach, H.Lehrach, P.Wincker, R.Reinhardt, and D.Chourrout. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature* **431**: 67-71.

Sessa,L., A.Breiling, G.Lavorgna, L.Silvestri, G.Casari, and V.Orlando. 2007. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA.* **13**: 223-239.

Shao,Z., F.Raible, R.Mollaaghababa, J.R.Guyon, C.T.Wu, W.Bender, and R.E.Kingston. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37-46.

Shaw,G., S.Morse, M.Ararat, and F.L.Graham. 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.* **16**: 869-871.

Shen,J., M.D.Person, J.Zhu, J.L.Abbruzzese, and D.Li. 2004. Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. *Cancer Res.* **64**: 9018-9026.

Shi,Y., F.Lan, C.Matson, P.Mulligan, J.R.Whetstine, P.A.Cole, R.A.Casero, and Y.Shi. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941-953.

Shiina,T., A.Ando, Y.Suto, F.Kasai, A.Shigenari, N.Takishima, E.Kikkawa, K.Iwata, Y.Kuwano, Y.Kitamura, Y.Matsuzawa, K.Sano, M.Nogami, H.Kawata, S.Li, Y.Fukuzumi, M.Yamazaki, H.Tashiro, G.Tamiya, A.Kohda, K.Okumura, T.Ikemura, E.Soeda, N.Mizuki, M.Kimura, S.Bahram, and H.Inoko. 2001. Genomic anatomy of a premier major histocompatibility complex paralogous region on chromosome 1q21-q22. *Genome Res.* **11**: 789-802.

Shogren-Knaak,M., H.Ishii, J.M.Sun, M.J.Pazin, J.R.Davie, and C.L.Peterson. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**: 844-847.

Shopland,L.S., C.R.Lynch, K.A.Peterson, K.Thornton, N.Kepper, J.Hase, S.Stein, S.Vincent, K.R.Molloy, G.Kreth, C.Cremer, C.J.Bult, and T.P.O'Brien. 2006. Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J. Cell Biol.* **174**: 27-38.

Shykind,B.M. 2005. Regulation of odorant receptors: one allele at a time. *Hum. Mol. Genet.* **14 Spec No 1**: R33-R39.

Simonis,M., P.Klous, E.Splinter, Y.Moshkin, R.Willemsen, W.E.de, S.B.van, and L.W.de. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**: 1348-1354.

Singh,P.B., J.R.Miller, J.Pearce, R.Kothary, R.D.Burton, R.Paro, T.C.James, and S.J.Gaunt. 1991. A sequence motif found in a Drosophila heterochromatin protein is conserved in animals and plants. *Nucleic Acids Res.* **19**: 789-794.

Sirotkin,A.M., W.Edelmann, G.Cheng, A.Klein-Szanto, R.Kucherlapati, and A.I.Skoultchi. 1995. Mice develop normally without the H1(0) linker histone. *Proc. Natl. Acad. Sci. U. S. A* **92**: 6434-6438.

Smith,F.J., A.D.Irvine, A.Terron-Kwiatkowski, A.Sandilands, L.E.Campbell, Y.Zhao, H.Liao, A.T.Evans, D.R.Goudie, S.Lewis-Jones, G.Arseculeratne, C.S.Munro, A.Sergeant, G.O'Regan, S.J.Bale, J.G.Compton, J.J.DiGiovanna, R.B.Presland, P.Fleckman, and W.H.McLean. 2006. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat. Genet.* **38**: 337-342.

Solomon,M.J., P.L.Larsen, and A.Varshavsky. 1988. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**: 937-947.

Solomon,M.J. and A.Varshavsky. 1985. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc. Natl. Acad. Sci. U. S. A* **82**: 6470-6474.

Southern,E., K.Mir, and M.Shchepinov. 1999. Molecular interactions on microarrays. *Nat. Genet.* **21**: 5-9.

Spector,D.L. 2003. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**: 573-608.

Spencer,V.A. and J.R.Davie. 2001. Dynamically acetylated histone association with transcriptionally active and competent genes in the avian adult beta-globin gene domain. *J. Biol. Chem.* **276**: 34810-34815.

Spitz,F., F.Gonzalez, and D.Duboule. 2003. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**: 405-417.

Sproul,D., N.Gilbert, and W.A.Bickmore. 2005. The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6**: 775-781.

Staege,M.S., C.Hutter, I.Neumann, S.Foja, U.E.Hattenhorst, G.Hansen, D.Afar, and S.E.Burdach. 2004. DNA microarrays reveal relationship of Ewing family tumors to both endothelial and fetal neural crest-derived cells and define novel targets. *Cancer Res.* **64**: 8213-8221.

Stein,A. and M.Mitchell. 1988. Generation of different nucleosome spacing periodicities in vitro. Possible origin of cell type specificity. *J. Mol. Biol.* **203**: 1029-1043.

Steinert,P.M., E.Candi, T.Kartasova, and L.Marekov. 1998a. Small proline-rich proteins are cross-bridging proteins in the cornified cell envelopes of stratified squamous epithelia. *J. Struct. Biol.* **122**: 76-85.

Steinert,P.M., T.Kartasova, and L.N.Marekov. 1998b. Biochemical evidence that small proline-rich proteins and trichohyalin function in epithelia by modulation of the biomechanical properties of their cornified cell envelopes. *J. Biol. Chem.* **273**: 11758-11769.

Sterner,R., G.Vidali, and V.G.Allfrey. 1979. Studies of acetylation and deacetylation in high mobility group proteins. Identification of the sites of acetylation in HMG-1. *J. Biol. Chem.* **254**: 11577-11583.

Stoler,S., K.C.Keith, K.E.Curnick, and M.Fitzgerald-Hayes. 1995. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev.* **9**: 573-586.

Strahl,B.D., R.Ohba, R.G.Cook, and C.D.Allis. 1999. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc. Natl. Acad. Sci. U. S. A* **96**: 14967-14972.

Su,A.I., M.P.Cooke, K.A.Ching, Y.Hakak, J.R.Walker, T.Wiltshire, A.P.Orth, R.G.Vega, L.M.Sapinoso, A.Moqrich, A.Patapoutian, G.M.Hampton, P.G.Schultz, and J.B.Hogenesch. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. U. S. A* **99**: 4465-4470.

Sullivan,K.F., M.Hechenberger, and K.Masri. 1994. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* **127**: 581-592.

Sumner,A.T. *Chromosomes: Organization and Function*.

Suto,R.K., M.J.Clarkson, D.J.Tremethick, and K.Luger. 2000. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* **7**: 1121-1124.

Swaminathan,J., E.M.Baxter, and V.G.Corces. 2005. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin. *Genes Dev.* **19**: 65-76.

Sybert,V.P., B.A.Dale, and K.A.Holbrook. 1985. Ichthyosis vulgaris: identification of a defect in synthesis of filaggrin correlated with an absence of keratohyaline granules. *J. Invest Dermatol.* **84**: 191-194.

Tagami,H., D.Ray-Gallet, G.Almouzni, and Y.Nakatani. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* **116**: 51-61.

Tanabe,H., F.A.Habermann, I.Solovei, M.Cremer, and T.Cremer. 2002a. Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutat. Res.* **504**: 37-45.

Tanabe,H., S.Muller, M.Neusser, H.J.von, E.Calcagno, M.Cremer, I.Solovei, C.Cremer, and T.Cremer. 2002b. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci. U. S. A* **99**: 4424-4429.

Taunton,J., C.A.Hassig, and S.L.Schreiber. 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**: 408-411.

Taylor-Papadimitriou,J., P.Purkis, E.B.Lane, I.A.McKay, and S.E.Chang. 1982. Effects of SV40 transformation on the cytoskeleton and behavioural properties of human keratinocytes. *Cell Differ.* **11**: 169-180.

Thambirajah,A.A., D.Dryhurst, T.Ishibashi, A.Li, A.H.Maffey, and J.Ausio. 2006. H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation. *J. Biol. Chem.* **281**: 20036-20044.

Thoma,F. and T.Koller. 1977. Influence of histone H1 on chromatin structure. *Cell* **12**: 101-107.

Thoma,F., T.Koller, and A.Klug. 1979. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83**: 403-427.

Thomas,J.O., C.Rees, and J.T.Finch. 1992. Cooperative binding of the globular domains of histones H1 and H5 to DNA. *Nucleic Acids Res.* **20**: 187-194.

Thorey,I.S., J.Roth, J.Regenbogen, J.P.Halle, M.Bittner, T.Vogl, S.Kaesler, P.Bugnon, B.Reitmaier, S.Durka, A.Graf, M.Wockner, N.Rieger, A.Konstantinow, E.Wolf, A.Goppelt, and S.Werner. 2001. The Ca2+-binding proteins S100A8 and S100A9 are encoded by novel injury-regulated genes. *J. Biol. Chem.* **276**: 35818-35825.

Thurman,R.E., N.Day, W.S.Noble, and J.A.Stamatoyannopoulos. 2007. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.* **17**: 917-927.

Tiwari,S., S.Ramachandran, A.Bhattacharya, S.Bhattacharya, and R.Ramaswamy. 1997. Prediction of probable genes by Fourier analysis of genomic sequences. *Comput. Appl. Biosci.* **13**: 263-270.

Tolhuis,B., R.J.Palstra, E.Splinter, F.Grosveld, and L.W.de. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**: 1453-1465.

Travers,A. 1999. The location of the linker histone on the nucleosome. *Trends Biochem. Sci.* **24**: 4-7.

Trowsdale,J. 2002. The gentle art of gene arrangement: the meaning of gene clusters. *Genome Biol.* **3**: COMMENT2002.

Tse,C., T.Sera, A.P.Wolffe, and J.C.Hansen. 1998. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Mol. Cell Biol.* **18**: 4629-4638.

Tsuji,T. and A.J.Cox. 1977. Ultrastructural studies of nuclei and nucleoli in psoriatic epidermis, compared with those in normal epidermis. *J. Invest Dermatol.* **69**: 205-210.

Tsukada,Y., J.Fang, H.Erdjument-Bromage, M.E.Warren, C.H.Borchers, P.Tempst, and Y.Zhang. 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**: 811-816.

Tummala,R., R.A.Romano, E.Fuchs, and S.Sinha. 2003. Molecular cloning and characterization of AP-2 epsilon, a fifth member of the AP-2 family. *Gene* **321**: 93-102.

Turner,B.M. 1993. Decoding the nucleosome. *Cell* **75**: 5-8.

Turner,B.M., A.J.Birley, and J.Lavender. 1992. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell* **69**: 375-384.

Vakoc,C.R., D.L.Letting, N.Gheldof, T.Sawado, M.A.Bender, M.Groudine, M.J.Weiss, J.Dekker, and G.A.Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**: 453-462.

van Attikum,H. and S.M.Gasser. 2005. The histone code at DNA breaks: a guide to repair? *Nat. Rev. Mol. Cell Biol.* **6**: 757-765.

van den Engh,G.J., R.Sachs, and B.J.Trask. 1992. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* **257**: 1410-1412.

van den Ijssel,P., M.Tijssen, S.F.Chin, P.Eijk, B.Carvalho, E.Hopmans, H.Holstege, D.K.Bangarusamy, J.Jonkers, G.A.Meijer, C.Caldas, and B.Ylstra. 2005. Human and mouse oligonucleotide-based array CGH. *Nucleic Acids Res.* **33**: e192.

van Holde,K. and J.Zlatanova. 1999. The nucleosome core particle: does it have structural and physiologic relevance? *Bioessays* **21**: 776-780.

Vermeulen,M., K.W.Mulder, S.Denissov, W.W.Pijnappel, F.M.van Schaik, R.A.Varier, M.P.Baltissen, H.G.Stunnenberg, M.Mann, and H.T.Timmers. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**: 58-69.

Vernet,G., M.Sala-Rovira, M.Maeder, F.Jacques, and M.Herzog. 1990. Basic nuclear proteins of the histone-less eukaryote Crypthecodinium cohnii (Pyrrhophyta): two-dimensional electrophoresis and DNA-binding properties. *Biochim. Biophys. Acta* **1048**: 281-289.

Verschure,P.J., d.K.van, I, L.W.de, d.van, V, A.E.Carpenter, A.S.Belmont, and D.R.van. 2005. In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Mol. Cell Biol.* **25**: 4552-4564.

Verschure,P.J., d.K.van, I, E.M.Manders, D.Hoogstraten, A.B.Houtsmuller, and D.R.van. 2003. Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep.* **4**: 861-866.

Versteeg,R., B.D.van Schaik, M.F.van Batenburg, M.Roos, R.Monajemi, H.Caron, H.J.Bussemaker, and A.H.van Kampen. 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* **13**: 1998-2004.

Volpi,E.V., E.Chevret, T.Jones, R.Vatcheva, J.Williamson, S.Beck, R.D.Campbell, M.Goldsworthy, S.H.Powis, J.Ragoussis, J.Trowsdale, and D.Sheer. 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113 (Pt 9)**: 1565-1576.

Wang,J., J.Mager, Y.Chen, E.Schneider, J.C.Cross, A.Nagy, and T.Magnuson. 2001. Imprinted X inactivation maintained by a mouse Polycomb group gene. *Nat. Genet.* **28**: 371-375.

Wang,Y., J.Wysocka, J.Sayegh, Y.H.Lee, J.R.Perlin, L.Leonelli, L.S.Sonbuchner, C.H.McDonald, R.G.Cook, Y.Dou, R.G.Roeder, S.Clarke, M.R.Stallcup, C.D.Allis, and S.A.Coonrod. 2004. Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science* **306**: 279-283.

Wang,Z.F., A.M.Sirotkin, G.M.Buchold, A.I.Skoultchi, and W.F.Marzluff. 1997. The mouse histone H1 genes: gene organization and differential regulation. *J. Mol. Biol.* **271**: 124-138.

Wei,Y., L.Yu, J.Bowen, M.A.Gorovsky, and C.D.Allis. 1999. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell* **97**: 99-109.

Weintraub,H. and M.Groudine. 1976. Chromosomal subunits in active genes have an altered conformation. *Science* **193**: 848-856.

Wells,R.S. and C.B.Kerr. 1966. Clinical features of autosomal dominant and sex-linked ichthyosis in an English population. *British Medical Journal* **1**: 947-950.

Widlund,H.R., H.Cao, S.Simonsson, E.Magnusson, T.Simonsson, P.E.Nielsen, J.D.Kahn, D.M.Crothers, and M.Kubista. 1997. Identification and characterization of genomic nucleosome-positioning sequences. *J. Mol. Biol.* **267**: 807-817.

Widom,J. 1996. Short-range order in two eukaryotic genomes: relation to chromosome structure. *J. Mol. Biol.* **259**: 579-588.

Wilkie,D. 1980. Pictorial Representation of Kendalls Rank Correlation Coefficient. *Teaching Statistics* **2**: 76-78.

Williams,B.A., R.M.Gwirtz, and B.J.Wold. 2006. Genomic DNA as a general cohybridization standard for ratiometric microarrays. *Methods Enzymol.* **410**: 237-279.

Williams,R.R., S.Broad, D.Sheer, and J.Ragoussis. 2002. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp. Cell Res.* **272**: 163-175.

Wirbelauer,C., O.Bell, and D.Schubeler. 2005. Variant histone H3.3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes Dev.* **19**: 1761-1766.

Wisniewski,J.R., A.Zougman, S.Kruger, and M.Mann. 2007. Mass spectrometric mapping of linker histone H1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Mol. Cell Proteomics.* **6**: 72-87.

Wolffe,A. *Chromatin: Structure and Function (3rd Edition)*.

Woodcock,C.L. 1994. Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *J. Cell Biol.* **125**: 11-19.

Woodcock,C.L., L.L.Frado, and J.B.Rattner. 1984. The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* **99**: 42-52.

Woodcock,C.L., S.A.Grigoryev, R.A.Horowitz, and N.Whitaker. 1993. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl. Acad. Sci. U. S. A* **90**: 9021-9025.

Woodfine,K., H.Fiegler, D.M.Beare, J.E.Collins, O.T.McCann, B.D.Young, S.Debernardi, R.Mott, I.Dunham, and N.P.Carter. 2004. Replication timing of the human genome. *Hum. Mol. Genet.* **13**: 191-202.

Wu,R.S. and W.M.Bonner. 1981. Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. *Cell* **27**: 321-330.

Wu,R.S., S.Tsai, and W.M.Bonner. 1982. Patterns of histone variant synthesis can distinguish G0 from G1 cells. *Cell* **31**: 367-374.

Wurtele,H. and P.Chartrand. 2006. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome. Res.* **14**: 477-495.

Yaar,M., M.S.Eller, J.Bhawan, D.D.Harkness, P.J.DiBenedetto, and B.A.Gilchrest. 1995. In vivo and in vitro SPRR1 gene expression in normal and malignant keratinocytes. *Exp. Cell Res.* **217**: 217-226.

Yang,Y.H., S.Dudoit, P.Luu, D.M.Lin, V.Peng, J.Ngai, and T.P.Speed. 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* **30**: e15.

Yokota,H., M.J.Singer, G.J.van den Engh, and B.J.Trask. 1997. Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome. Res.* **5**: 157-166.

Yokota,H., E.G.van den, J.E.Hearst, R.K.Sachs, and B.J.Trask. 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J. Cell Biol.* **130**: 1239-1249.

Yuan,G.C., Y.J.Liu, M.F.Dion, M.D.Slack, L.F.Wu, S.J.Altschuler, and O.J.Rando. 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* **309**: 626-630.

Zakany,J., M.Kmita, and D.Duboule. 2004. A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* **304**: 1669-1672.

Zappavigna,V., L.Falciola, M.Helmer-Citterich, F.Mavilio, and M.E.Bianchi. 1996. HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. *EMBO J.* **15**: 4981-4991.

Zhang,D.E. and D.A.Nelson. 1988. Histone acetylation in chicken erythrocytes. Rates of deacetylation in immature and mature red blood cells. *Biochem. J.* **250**: 241-245.

Zhang,H., D.N.Roberts, and B.R.Cairns. 2005. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**: 219-231.

Zhang,Y. and D.Reinberg. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**: 2343-2360.

Zhao,X.P. and J.T.Elder. 1997. Positional cloning of novel skin-specific genes from the human epidermal differentiation complex. *Genomics* **45**: 250-258.

Zhao,Y., A.Terron-Kwiatkowski, H.Liao, S.P.Lee, M.H.Allen, P.R.Hull, L.E.Campbell, R.C.Trembath, F.Capon, C.E.Griffiths, D.Burden, R.McManus, R.Hughes, B.Kirby, S.F.Rogers, O.Fitzgerald, D.Kane, J.N.Barker, C.N.Palmer, A.D.Irvine, and W.H.McLean. 2007. Filaggrin null alleles are not associated with psoriasis. *J. Invest Dermatol.* **127**: 1878-1882.

Zink,D., H.Bornfleth, A.Visser, C.Cremer, and T.Cremer. 1999. Organization of early and late replicating DNA in human chromosome territories. *Exp. Cell Res.* **247**: 176-188.

Zink,D., T.Cremer, R.Saffrich, R.Fischer, M.F.Trendelenburg, W.Ansorge, and E.H.Stelzer. 1998. Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.* **102**: 241-251.

Zink,D. and R.Paro. 1995. Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. *EMBO J.* **14**: 5660-5671.

Zinner,R., H.Albiez, J.Walter, A.H.Peters, T.Cremer, and M.Cremer. 2006. Histone lysine methylation patterns in human cell types are arranged in distinct three-dimensional nuclear zones. *Histochem. Cell Biol.* **125**: 3-19.

Zirbel,R.M., U.R.Mathieu, A.Kurz, T.Cremer, and P.Lichter. 1993. Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome. Res.* **1**: 93-106.

Zlatanova,J. and D.Doenecke. 1994. Histone H1 zero: a major player in cell differentiation? *FASEB J.* **8**: 1260-1268.

Zlatanova,J. and K.van Holde. 1998. Linker histones versus HMG1/2: a struggle for dominance? *Bioessays* **20**: 584-588.

Zorn,C., C.Cremer, T.Cremer, and J.Zimmer. 1979. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp. Cell Res.* **124**: 111-119.

Zorn,C., T.Cremer, C.Cremer, and J.Zimmer. 1976. Laser UV microirradiation of interphase nuclei and post-treatment with caffeine. A new approach to establish the arrangement of interphase chromosomes. *Hum. Genet.* **35**: 83-89.

Zwilling,S., H.Konig, and T.Wirth. 1995. High mobility group protein 2 functionally interacts with the POU domains of octamer transcription factors. *EMBO J.* **14**: 1198-1208.